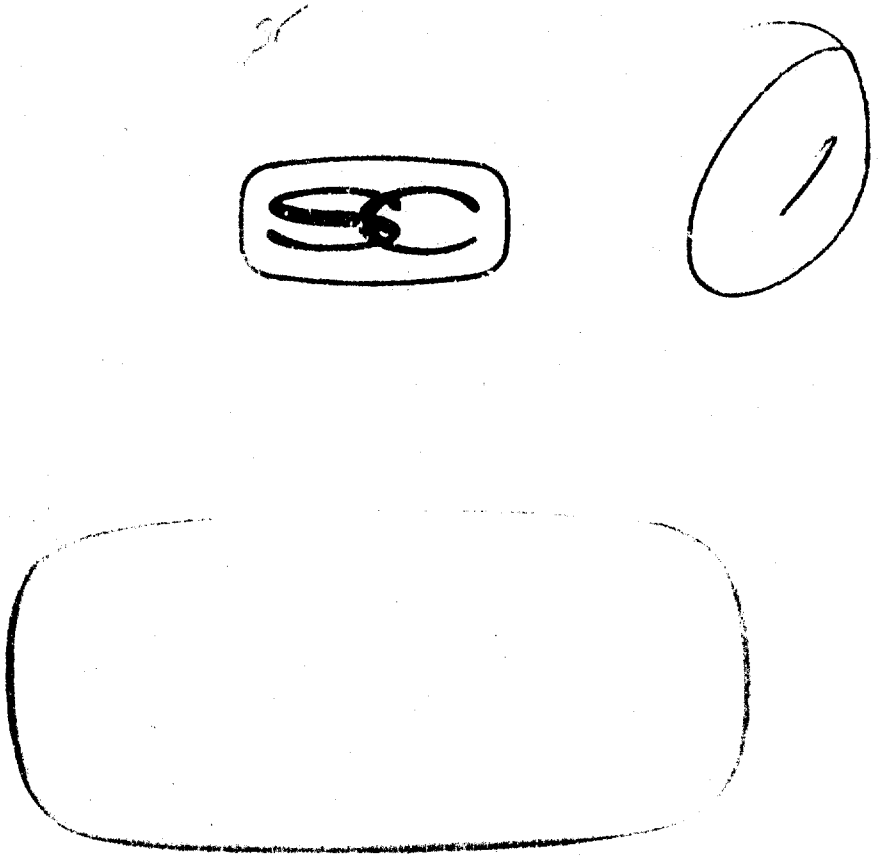
AD 652742



Best Available Copy

DDC
RETRIEVAL
JUN 9 196
E

STATEMENT NO. 1
Distribution of This Document is Unlimited

ARCHIVE COPY

Science Communication
INCORPORATED
1079 Wisconsin Avenue, N.W.
Washington, D. C.

# Chemical Data Compilation Analysis Survey

Contract NSF C-478
20 March 1967

R. G. Brunner, B. K. Farris, S. I. Jover, and D. O. Myatt

Conducted for

Chemical Information Unit
Office of Science Information
National Science Foundation

This report was prepared under auspices of the two-year Interagency Chemical Information Program, which is co-sponsored by the Department of Defense, the National Institutes of Health, and the National Science Foundation.

by

SCIENCE COMMUNICATION, INC.
1081 Wisconsin Avenue, N. W.
Washington, D. C. 200 07

**Science Communication**
Washington, D. C. 200 07

NSF C-478 Final Report
20 March 1967

Table of Contents

**Science Communication**
Washington, D. C. 200 07



## FOREWORD

The generation, validation, and correlation of numerical data is widely recognized as a centrally important aspect of scientific and technical activity. The technical community exploring and exploiting chemical knowledge has taken a leading position in developing the collation, evaluation, correlation and extrapolation techniques required for use of large sets of interrelated numerical data. The content and format of these chemical data* compilations reflect a considered effort to meet the information communication requirements of specific user populations.

The complexity of the present chemical information-communication matrix is difficult to describe analytically. Projection of a more effective future system from this non-quantitative reference base is obviously even more difficult. However, numerical chemical data compilations existing in the matrix are amenable to systematic characterization. Thus, they can facilitate a more quantitative definition of existing chemical data communication systems and, subsequently, criteria for improved systems.

Data compilations are technically and economically viable in actual chemical communication situations. It appears axiomatic that the chemical data and information systems of the future must be based on an awareness of the capabilities and limitations of present practices and resources.

It is widely recognized that the availability of high capacity automatic data processing and transmission equipment has opened new opportunities for the construction and integration of files of data on chemical substances. Currently, large scale systems are under consideration for handling chemical information and data. These include the Chemical Information Program coordinated by the National Science Foundation, the Chemical Information and Data System development program of the Department of the Army, and the National Standard Reference Data System directed by the National Bureau of Standards. Each of these programs includes efforts toward the development of methods of analyzing data files and data flows to obtain information needed to plan improved systems.

---

* In the context of this report, chemical data is defined as quantitative information concerning the physical and chemical properties of chemically identifiable substances or systems.

Chemical data handling by automated equipment, to date, has paralleled early business data handling in that it has been pursued in a fragmented manner. During the early phases of applying automated equipment to the handling of business data, separate files were often maintained to service a limited set of user requirements. Other files containing much of the same data were maintained to serve other user requirements, resulting in fragmentation of the total data file. More recently, integrated data processing, which combines files, eliminates or reduces duplication of data items, and serves all users from fewer data files, has proven more efficient in business data processing. Conditions contributing to this change in business data processing included:

- Development of computers with larger storage memories and faster processing capabilities, and
- Development of methods of analyzing data files and data flows to obtain information needed to plan improved systems.

The applicability of integrated data processing techniques to chemical data handling and user servicing is largely an unresolved question. A major objective of NSF Contract C-478 is the development and application of a methodology for analysis and characterization of chemical data compilation activities. The basic analysis technique which we tested was the examination of each data compilation as a data file and its substance and property coverage. However, because of the complex interactions which exist between a data compilation activity and the environment on which it is operated and the technical activity which it supports, our study also examined operating statistics, operating motives, users served and other factors that might shed light on basic design factors.

## I. Introduction

The National Science Foundation, which is sponsoring the National Chemical Information Program, contracted with Science Communication, Inc., under contract number NSF C-478, for assistance in analyzing existing compilations of data on the physical and chemical properties of chemical substances and systems. Our work scope was initiated as a study of the structure of existing chemical data compilations, but due to new developments in the National Chemical Information Program, the emphasis was shifted towards obtaining a greater amount of information relevant to the design elements associated with the actual development of a national chemical information system. Our activities consequently, were phased with those of the Chemical Abstracts Service in the development of a chemical compound registry system, and with system pre-design studies currently under way for the National Science Foundation by Information Management, Inc.

Prior to discussion of the findings, it would be desirable to place our study effort in its context among the several existing modes of communication of chemical data from generator to user.

A community of interest exists when individuals and/or organizations identify themselves with a common goal. In scientific and technical endeavors, such goals are often referred to as missions. In this context, missions are not necessarily applications-oriented. One normal manifestation of a community of interest is the development of an effort to generalize, organize and conserve the data required to pursue the mission. This effort may be well-articulated and formally structured, or it may be informally structured and hardly discernible. Various communities of interest have developed chemical data communication systems which operate in one or more of the following linkages:

Generator to user
Generator to document publisher to user
Generator to document publisher to document processor to user
Generator to document publisher to document processor to data processor to user
Generator to data processor to user
Generator to data processor to document processor to user

Our study effort was restricted to the consideration of the data processor element of the system and the interfaces between the data processor and the elements which provide input or receive output. Input and output elements for the data processor elements depend on the operating mode of a system. In many systems, a data processor (or, in our project terminology, a data compilation activity) may operate simultaneously as an element of more than one system. That is, a data compilation activity often performs several communication functions simultaneously.

## II. Study Procedure

We have studied a significant number of data compilations and their respective centers for obtaining quantitative information representative of their current activity. Parameters studied include those representing scientifically meaningful parcels of information as well as the fundamental elements of such parcels. Data of the following nature were collected or generated on selected data compilations:

(1) Coverage of chemical substances including types of chemical substances (pure chemicals or mixtures, elements or compounds, inorganic or organic compounds, etc.).

(2) Mode of representation of chemical substances (CA name, molecular formula, structural formula, code, etc.).

(3) Number and identity of properties covered.

(4) Format of data compilation.

(5) Parameters or units of measure of the property.

(6) Number of associated variables in the file record.

(7) Reference citing.

(8) Extent of validation or evaluation of the property values presented in the output.

(9) Available forms of the compilation (printed handbook, computer tapes, optical coincidence indexes, tabular indexes, edge-notched cards, 3 x 5 cards, etc.).

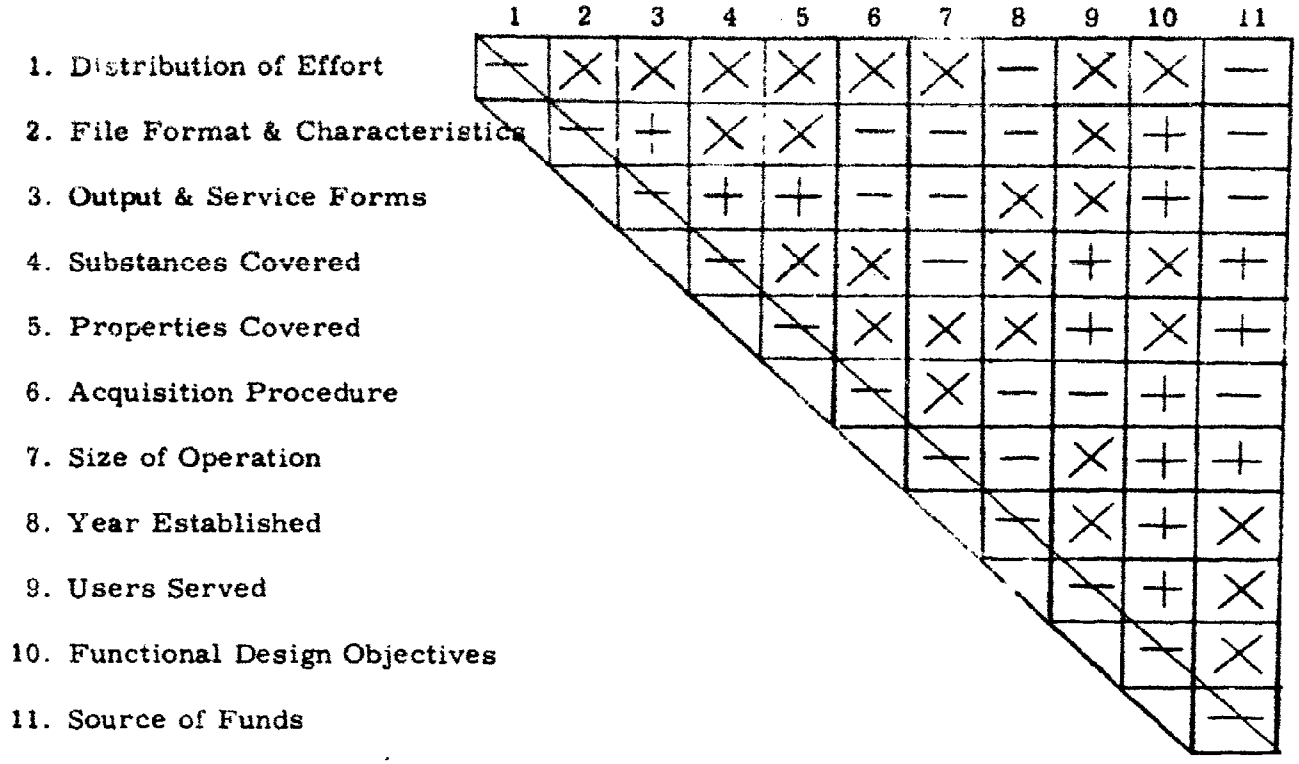
(10) Time period coverage and growth rate of compilation.

(11) File structures developed to accommodate the above.

Data for our analysis were first obtained through a review of and extraction from specimen reports and descriptive literature issued by data compiling activities. Considerable amounts of time and effort were spent in obtaining this data from sources in the SCI files, the National Standard Reference Data Office, the NAS-NRC Office of Critical Tables, etc.

A second step of our analysis was to obtain, through interviews with data compilation directors and personnel, verification of our preliminary findings, operating statistics, communities the compilation is intended to serve, current design rationales, experience with other formats, and information we had not acquired from our review of their descriptive literature and specimen outputs. As much information and data as possible was extracted and placed on work sheet forms before the data compilation director was interviewed. By using this approach interviews consisted of informed questions covering the internal and external operations of the activity. Interviews were productive, and cooperation was excellent.

Analysis of the collected information began as soon as meaningful amounts had been obtained. In the course of the development of our information gathering worksheets, we provided for the principal correlations which we intended to perform with the assembled information. These correlations are identified in the following matrix (X indicates performance of a correlation; -- indicates no correlation; the more important correlations to be performed are indicated by +).




Correlation Matrix

NOTE: The matrix row numbers are the same as the column numbers.

The correlation objectives place particular emphasis on:

(1) identification and characterization of significant information used in chemical data processing,

(2) realistic estimates of the volumes of chemical data compiled to serve identified user needs,

(3) identification of substance and property coverage among chemical data compilations, and

(4) the implications of the structure and content of existing chemical data compilations for design and operation of future automated chemical data and information systems.

## III. Work Accomplishments

Our initial efforts were directed toward the development of criteria for the selection of data compilation activities to be analyzed. The basic criteria established for the data compilations studied are that they should:

(1) cover either physical or chemical properties of chemically identifiable substances,

(2) cover a significant segment of chemical substances or properties,

(3) serve a significant user population,

(4) disseminate compilations of data to the technical community, and

(5) have existed on a continuing basis for a period adequate to verify operational validity of use to the development of a National Chemical Information System.

The fifth criterion was a screening consideration in our selection of data compilation activities to be analyzed. The activities chosen for study fulfill all criteria in varying degrees, but the fifth criterion was in many cases a deciding factor for or against inclusion.

The office of Standard Reference Data of the National Bureau of Standards made available to us early in the project the preliminary tabulation of results from the following question in their survey of the data use practices of the members of the American Chemical Society:

> "Which data compilations do you consult most often? Please list small, highly specialized compilations as well as more extensive compendia."

As might be expected, the response to this survey showed that handbooks and large compendia such as Lange's Handbook of Chemistry or the International Critical Tables proved to be useful to a significantly large number of users. Significant though such collections are, however, in the realm of chemical data their modus operandi of periodic revision of an editorial nature did not meet the ~~fourth~~ 5th criterion sufficiently to merit coverage.

Efforts were made to include a representative sampling of each of the following chemical substances and systems:

(1) pure chemical elements,
(2) pure chemical compounds (organic and inorganic), and
(3) mixtures of compounds and/or elements.

A literature search was conducted to identify candidate chemical data compilation activities for analysis. A list of candidate compilations was developed through the examination of standard directories to technical information services and systems, and other local sources.

Efforts to characterize selected compilations in an analytical fashion revealed that compilation activities which were initially considered as a single compilation were in fact multiple compilations prepared by one housing institution. We therefore found it necessary to characterize the parameters of each compilation where several compilations were prepared in one housing institution. This provided an opportunity to look for parameters of a compilation that were influenced by the basic communication requirements, as distinguished from those conditioned by the housing environment (e.g., processing traditions, availability of special support facilities etc.). Other influences, such as the special skills of the center's key staff, the nature of the subject field, sponsor influence on the center's mission, and the limited number of housing institutions in the sample swamped out any distinctions that could be identified with housing institutions.

Data compilation activities examined were in the following housing institutions:

(1) Oak Ridge National Laboratory, Atomic Energy Commission, Oak Ridge, Tennessee: **Atomic and Molecular Processes Information Center.**

(2) National Bureau of Standards, Boulder, Colorado: **Cryogenic Data Center.**

(3) Electrical-Electronic Properties Information Program, Hughes Aircraft Company, Culver City, California: **Electronic Properties Information Center.**

(4) Dow Chemical Company, Midland, Michigan: **Joint Army-Navy-Air Force (JANAF) Thermochemical Tables.**

(5) Texas A&M University, College Station, Texas: **Thermodynamics Research Center Projects on Properties of Chemical Compounds** (API Project and TRC, formerly MCA).

(6) National Bureau of Standards, Gaithersburg, Maryland: **Selected Values of Thermodynamic Properties of Inorganic Substances Data Center** (NBS Circular 500).

(7) University of California, Berkeley, California: **Thermodynamic Properties of Metals and Alloys Data Center.**

(8) Purdue University, Lafayette, Indiana: **Thermophysical Properties Research Center.**

(9) Wyandotte Chemical Corporation, Wyandotte, Michigan: **The Wyandotte-ASTM Spectral Data Project.**

(10) American Society for Testing and Materials, Philadelphia, Pennsylvania: **The Joint Committee on X-Ray Powder Diffraction Data.**

The information collected for analysis was grouped into eleven categories. The primary and secondary means of acquiring each category type of information were formulated as shown in the following table:

Sources of Information for Analysis

| Information | Primary Source | Secondary Source |
|---|---|---|
| (1) Functional design objectives | Interview | Descriptive Literature |
| (2) Users served | Interview | Descriptive Literature |
| (3) Substances covered | Specimen analysis | Descriptive Literature |
| (4) Properties covered | Specimen analysis | Descriptive Literature |
| (5) Data acquisition procedures | Interview | Descriptive Literature |
| (6) File formats and characteristics | Specimen analysis | Interview |
| (7) Output and service forms | Specimen analysis | Interview |
| (8) Size of activity | Descriptive literature | Interview |
| (9) Distribution of effort | Interview | Descriptive Literature |
| (10) Source of Funds | Interview | Descriptive Literature |
| (11) Year established | Descriptive literature | Interview |

In accordance with the above plan, work sheets were designed and developed, and a file of descriptive literature for each compilation to be studied was assembled. The descriptive literature file was assembled from the SCI library and its bibliographic references, the Defense Documentation Center, Technical Abstracts Bulletin, and direct inquiries to the compilation activities. Annual reports, progress reports, and their discussions of coverage, processing methods, operating statistics, and services rendered proved to be the most useful type of descriptive literature. Our initial worksheets were developed for gathering of information from such literature and from specimen analysis. Information gathered with this worksheet was adequate for a general characterization of each compilation activity, but was not sufficiently definitive to characterize quantitatively a specific compilation prepared in an activity which produces several compilations. We also encountered problems resulting from the heterogeneity of terminology found in the descriptions of compilation activity, and it became apparent that this heterogeneity would at some time have to be resolved by a generalized vocabulary or concept of data compilation operations if a systematic characterization and cross-section analysis of different compilations were to be accomplished. We, therefore, decided to develop a generalized concept of data compilation operations and to incorporate it into the information-gathering worksheet, so that information would be uniformly gathered and the problem resolved in the course of our work, instead of being an obstruction to its completion.

The following table identifies typical processing operations we identified as basic to data compilation activities:

Typical Processing Operations in Data Compilation Activities

A. Data Source Processing Operations

1. Search for source documents
2. Obtain source documents
3. Screen source documents for relevance and duplication
4. Catalog source documents
   a. bibliographic characterization
   b. content characterization
5. Process source documents for storage
6. Input source documents to storage
7. Search source documents
8. Output references from search of source document file

B. Data File Construction

1. Assemble appropriate source documents
2. Extract data and appropriate information
3. Screen and evaluate extracted data and information
4. Process data and information for storage

C. Output Operations

1. Interpret data inquiry for publication requirement and formulate appropriate search strategy
2. Search and retrieve data and information
3. Edit search output
4. Disseminate appropriate data and information

Generalized flow charts of these processes were developed (figures 1,2,3), and our information-gathering worksheet was redesigned to gather information in reference to this conceptualization of the data compilation process.

Experience proved that the flow charts and operations, with minor changes, satisfactorily covered the principal operations of data compilation activities of interest to the study. By incorporating this concept into the worksheet, information given on an individual level in progress and technical reports of the data centers was reduced to a uniform level. The concept proved satisfactory in interviews with the compilation directors. It eliminated confusion and aided communication between the interviewer and interviewee. Not all of the operations in the conceptual model were performed by all of the data compilation activities examined. Some of the operations and services were omitted entirely by some activities, performed simultaneously in a single operation, or were far more important in one compilation activity than in another. This variety was expected. The conceptualization, however, served its purpose in reducing such variety from descriptive paragraphs to specific terms and operational elements.

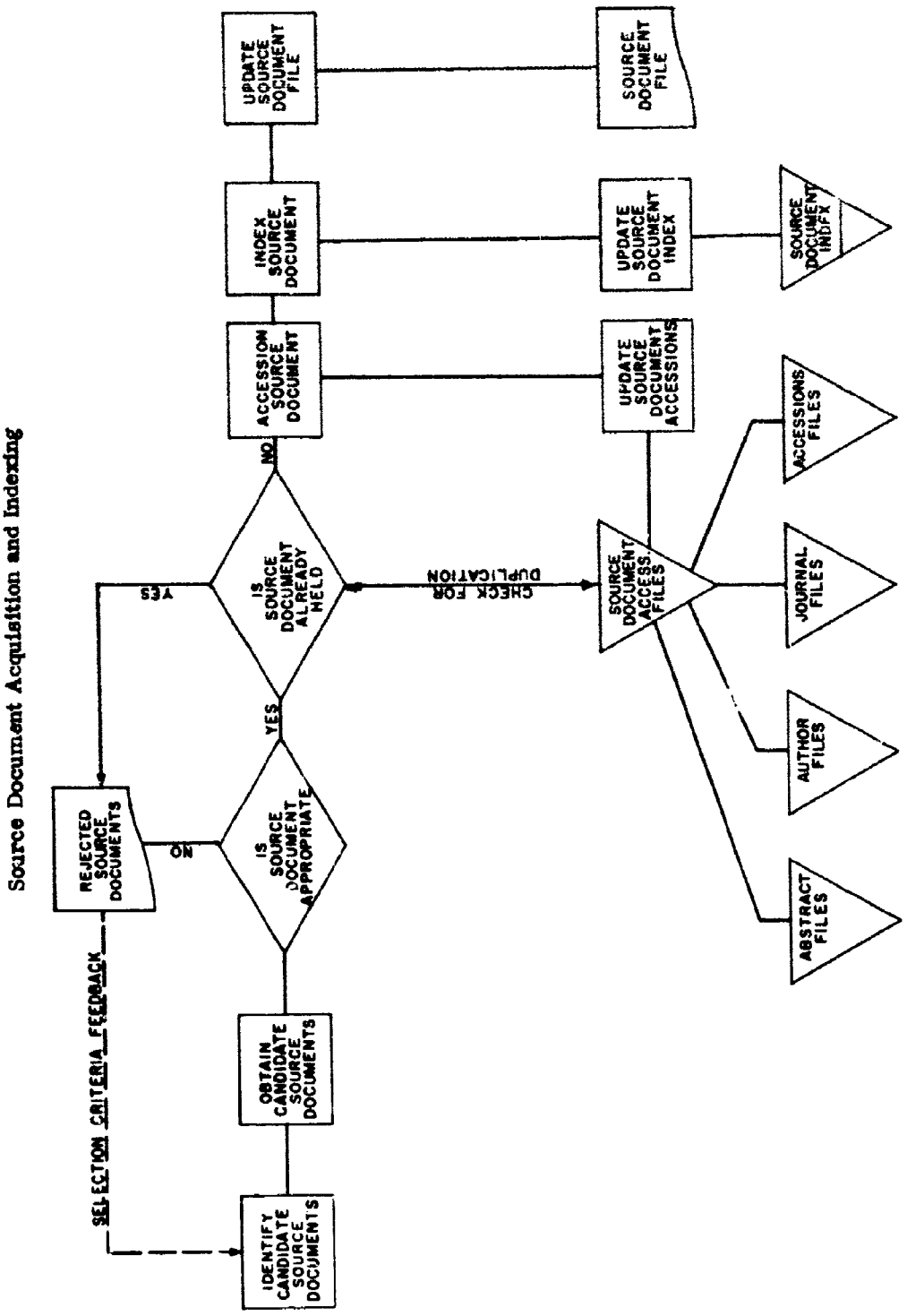



FIGURE 1.
Source Document Acquisition and Indexing

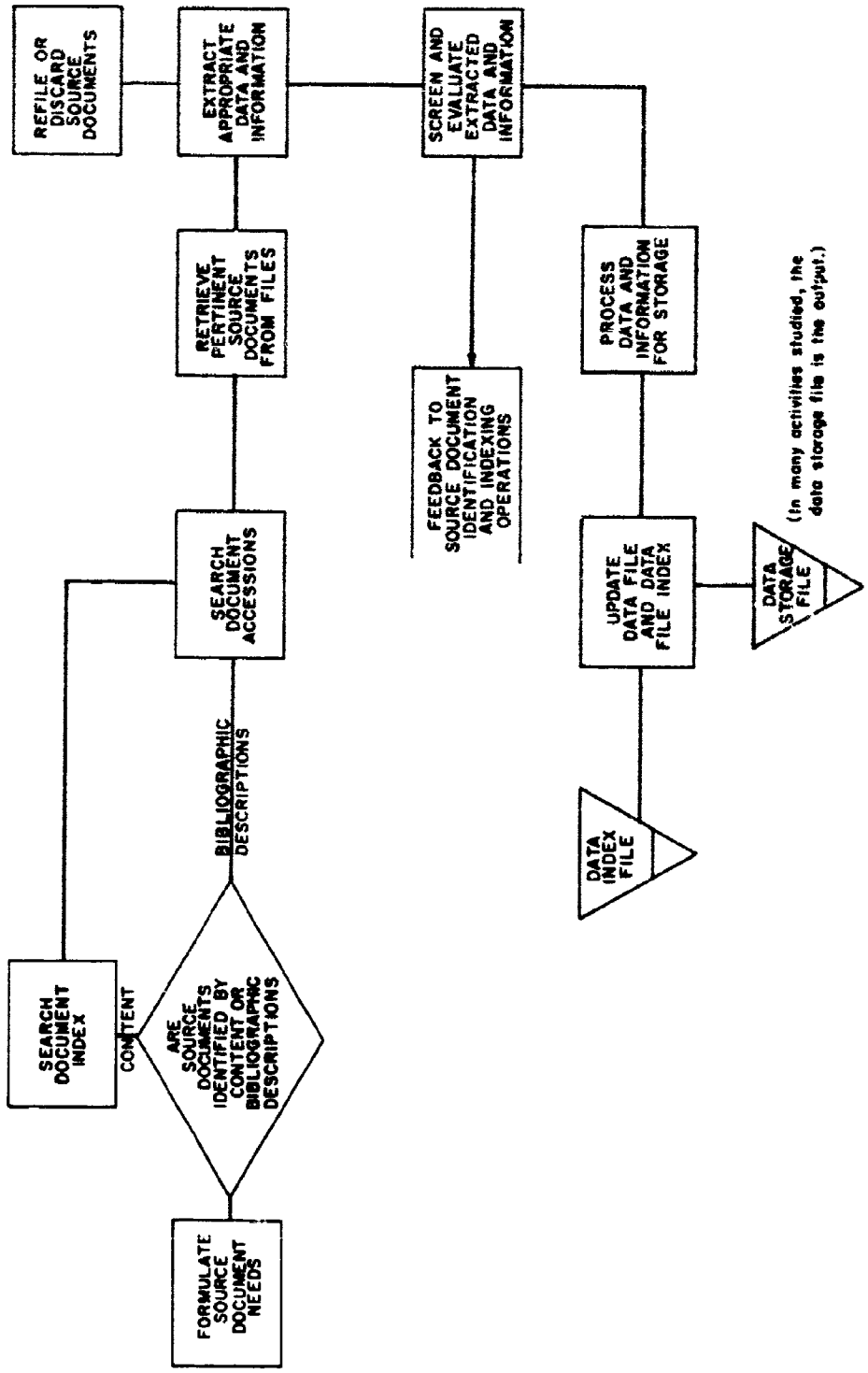



FIGURE 2.

Data File Construction

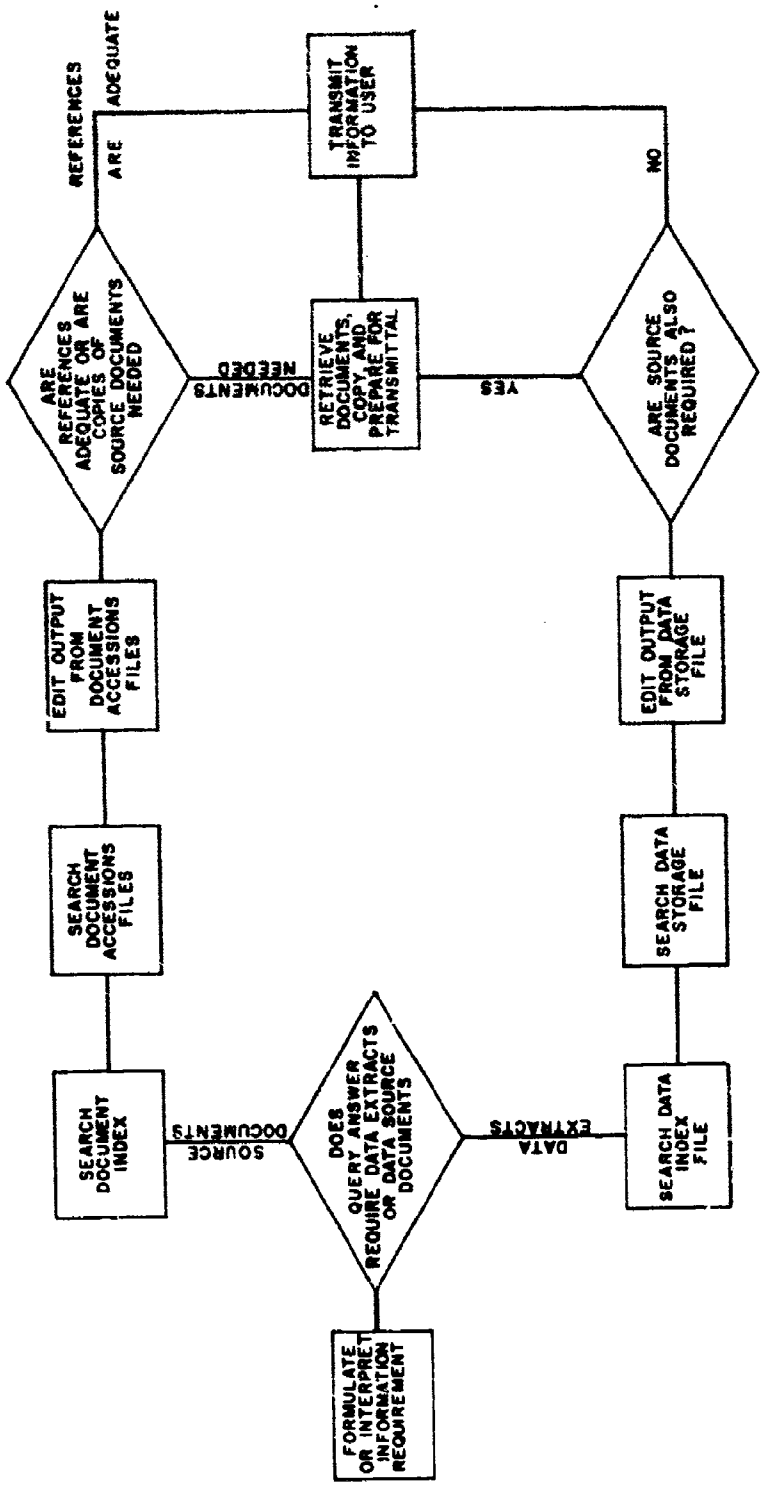



FIGURE 3.

User Servicing Operations

The table given below is a coarse description that highlights the principal operational file elements of the compilation activities studied.

| Center Code | Center Name | Data Source File | Data Source & Data File Combination | Data File |
|---|---|---|---|---|
| EPIC | Electronic Properties Information Center | X | -- | -- |
| TPRC | Thermophysical Properties Research Center | X | -- | -- |
| UC | Thermodynamic Properties of Metals and Alloys Data Center, U. Calif., Berkeley | X | -- | -- |
| CDC | Cryogenic Data Center | X | -- | -- |
| AMPIC | Atomic & Molecular Processes Information Center | X | -- | -- |
| A&M | Thermodynamics Research Center, Texas A&M | -- | X | -- |
| NBS | Selected Values of Thermodynamic Properties of Inorganic Substances (NBS Circular 500) | -- | X | -- |
| JANAF | Joint Army-Navy-Air Force Thermochemical Tables | -- | X | X |
| ASTM | Wyandotte-ASTM Spectral Data Project | -- | -- | X |
| X-Ray | Joint Committee on X-Ray Powder Diffraction Data | -- | -- | X |

The information-gathering worksheet was designed to acquire information on the technique and rationale of each compilation activity for each of the flow charted operations that it actually performs. The final version of the worksheet was highly structured (see Appendix D) and was used for acquisition of information from each of the sources given earlier. The information-gathering sequence of descriptive literature, specimen analysis, and interview proved effective.

## IV. Findings

Meaningful analyses for our objectives could be made on any of the following levels:

(1) Viewing each compilation activity historically, as it evolved to satisfy the specific user needs in its unique area of coverage and resolve its specific data handling problems;

(2) Analyzing all data compilation activities across the board to study the larger problems which they face in common;

(3) Analyzing groupings of compilations that were distinct and meaningful, either through similarity in the types of data which they handle, in compilation formats, or in the various services they provide.

All three levels proved useful in the analysis.

Compilation Center Activities. A listing of the data compilations produced by the centers interrogated is given in Appendix B. Appendix C contains a summary description of each activity's objectives and operational procedures, flow charts of its document- and data-processing operations for compilation and output development, and a chart describing the physical arrangement and content of its principal internal files for accomplishing these operations. Figure 4 presents groupings of data compilations in terms of output and coverage classes that were considered best suited for more detailed analysis. The groups established are by no means mutually exclusive. They are used to delineate broad views of the variety of compilation activities we identified and to characterize the groups. Compilations presenting graphic data, which essentially provide an infinite number of values, do not furnish statistics compatible with compilations presenting single values. Likewise, compilations comprising cumulative tables in prescribed substance-property formats are amenable to quantitative analysis, while monographs required a descriptive level of analysis to accommodate varying formats of presentation, and interpretive discussions associated with the values given. Efforts were made to display these intrinsic differences in the master worksheet tabulation in Appendix A. The occasional "not applicables" in the tabulations indicate that this was not always possible.

Other significant distinctions beside those noted by Figure 4 exist among the data compilation activities. Although we found it necessary to construct the generalized model represented by the worksheet as a check-point sequence in our information-gathering operation, we did not expect actual data compilation activities to fit any single model. A second model built on the operating and service philosophies associated with a data compilation activity proved

Figure 4.

FORMATS AND TYPES OF DATA
IN SELECTED COMPILATIONS

| | Output Format | | Highly Eval Data | Mod. Eval Data | Chem Ident Mat. | Chem Unident Data | Graph Data Output | Num Data Output | Analyt Chem Data | Thermo-dyn-chem Data | Thermo-Physic. Data |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cont'g Tbls | Monog. | | | | | | | | | |
| Cryogenic | - | X | X | X | X | - | X | X | - | X | X |
| EPIC | - | X | - | X | X | X | X | X | - | - | electron. - electric. |
| JANAF | X | - | X | - | X | - | X | - | - | X | - |
| API - TRC (spect. data) | X | - | X | - | X | - | X | - | X | X | - |
| API - TRC (select values) | X | - | X | - | X | - | - | X | - | - | - |
| Circ. 500 - NBS | X | - | X | - | X | - | - | X | - | X | - |
| UC - Berkeley | X | - | X | X | X | - | X | X | - | X | - |
| Purdue - TPRC | X | - | X | X | X | X | X | X | - | - | X |
| Wyandotte - ASTM | X | - | - | X | X | X | - | X | X | - | - |
| X - Ray | X | - | X | - | X | X | - | X | X | - | - |
| Atom & Molec. Processes AMPIC | - | X | - | - | X | X | - | - | - | - | - |

to be useful in analysis of different patterns actually found, and should have value in recognizing the design requirements of future data compilation operations. This model comprised the following sets of servicing and operating objectives:

Data Acquisition and Screening Objective

- Collect and disseminate all data in area covered.
- Collect all data and disseminate only selected data in area covered.
- Collect and disseminate only selected data in area covered.

Information Provided Users

- References to sources of data.
- Usable data with reference to information sources.
- Usable data without reference to information sources.

Type of User Service

- Response to individual inquiries.
- Periodic dissemination of accumulated information or data.

Other Purposes of Compilation Activity

- Reduce costs for the user to obtain reference and/or information.
- Increase the utility of data by conversion to a more useful format.
- Generate new data by computation or extrapolations with data compiled.
- Conserve scientific manpower by eliminating duplication of evaluation efforts.

This model provides an insight as to how current compilation activities are designed and operated in order to meet the needs of identified users, and was incorporated into our information gathering worksheets and analysis procedures. The following findings resulted:

A. Data Acquisition and Screening

Three basic types of compilation activity were postulated in our model:

(1) Collect all data and output all data, with only moderate evaluation,

(2) Collect all data and output selected, highly evaluated data, and

(3) Collect selected data and output selected, highly evaluated data.

Activities limited to Category (1) are restricted essentially to accumulating and reformatting functions. Such activities are characterized by high output volumes. An example of this type of operation would be the Wyandotte-ASTM Spectral Data Project, which collects all published and commercial infrared and visible spectral curves. Wyandotte then codes the spectra for both band positions and chemical structure and places the coded data in a standardized format for dissemination into a variety of publications, which range from tabular indices to computer tapes.

The majority of data compilation groups in our sample fall in Category (2). These are chemical data compilation activities which screen the entire open literature for measured values, organize these sources of information, and then evaluate the data to produce a "best estimate," recommended value. The requirement to identify the original source document in the data evaluation process provides the raw material for a bibliographic referencing service, and most Category (2) activities offer such services.

None of the activities in our sample fell into Category (3).

As will be noted in Figure 5, some of the centers in our sample fall into more than one category. The compilation level can be influenced, for example, by the substances and properties for which data is given. JANAF, for example, presents property data for pure substances in an area where basic data and knowledge of substance-property relationships is sophisticated. They store evaluated fundamental thermochemical constants and mathematical formulae on magnetic tapes which they use to generate extensive tables of thermochemical functions. On the other hand, the Thermophysical Properties Research Center at Purdue University collects data on the thermo-transport properties of all substances, many of which are solids in which trace impurities and structural defects have major influence on properties. In this field, rigorous establishment of new values by computation or extrapolation from other values is difficult or impossible. Considerable empirical data and judgemental appraisal of reported measurements must be assembled before an expert can ascertain the most probable value for the purposes of identified user interests. Even then, TPRC has found it most valuable to disseminate in a concise format all the reliable empirical data used to ascertain these most probable values, because of the composition variances in the substances covered and similar variances in the alternate grades of a material available to the user.

The inverse relationship between the volume of selected values presented in the Thermophysical Properties Research Center Data Book, compared with the JANAF Thermochemical Tables, probably results largely from this difference in substance-property characterization. The TPRC Data Book inputs and displays a large mass of data with the final objective being to

FIGURE 5.

SERVICE PHILOSOPHIES AND EMPHASES

| | UC - Berkeley | Texas A & M | JANAF | NBS, Circular 500 | Purdue - TPRC | Cryogenic | X - Ray | Wyandotte - ASTM | EPIC | Atom & Molec. Processes |
|---|---|---|---|---|---|---|---|---|---|---|
| a. Collect all data and disseminate moderately evaluated data in area covered. | N | N | N | N | P | N | N | P | N | N |
| b. Collect all data and disseminate highly evaluated or selected data in area covered. | P | P | P | P | P | P | P | N | P | P |
| c. Collect and disseminate only selected data (e. g., most probable values) in area. | N | N | N | N | N | N | N | N | N | N |
| d. Provide user with references to sources of data. | P | S | S | S | P | P | P | P | P | P |
| e. Provide user with data and with information source references | P | P | N | P | P | P | P | P | P | P |
| f. Provide user with data without information source references. | N | N | P | P | N | P | P | N | N | P |
| g. Provision of quick response to individual queries. | N | N | D | D | P | S | S | S | P | D |
| h. Periodic dissemination of accumulated information or data. | P | P | P | P | P | P | P | P | P | P |
| i. Fast processing of newly available data. | D | D | D | D | D | D | P | D | D | S |
| j. Fast dissemination of newly available data. | D | D | S | D | D | D | P | D | D | S |
| k. Reduce costs for user to obtained references and/or data. | S | S | P | S | P | S | D | D | S | S |
| l. Increase utility of data by conversion to more useful format (e. g. graphic or ADP). | S | S | S | N | P | S | P | P | S | N |

P = A primary objective

S = A secondary objective

D = Desirable, but subordinated by other factors

N = Not a pertinent objective

output a single best estimate curve in the context of all relevant data. In contrast the JANAF Thermochemical Tables need only input and display a limited number of property values (which are documented on the reverse side of each page of tables) with the output resulting in an extensive table of thermochemical functions.

B. Information Provided User

The type of data processed is not the sole determinant for the category of data compilation output which will be established. The user of the output of the compilation is also a significant factor. For example, the design engineer is quite willing to accept a single best-estimate value for the heat transfer coefficient of aluminum; in contrast, a research scientist planning to remeasure this value would prefer to have access to all previously measured values.

One of the supplemental information outputs most frequently associated with the centers examined is a bibliographic reference to the source of the data. In addition, we found that data compilation activities normally exhibit an evolutionary development. In the history of individual centers, they often first provide data reference services, usually in the form of bibliographies indexed to identify the substance and property coverage of the references included. Data file construction and servicing of data needs of users appeared to evolve sequentially as the "store" became larger and the center extracted and organized sufficient amounts of data. Our study efforts included only those compilation activities which provide the user with some form of data. There are undoubtedly activities now in operation whose mission is to provide data services, but whose operational level is still at the bibliographic level.

Regardless of the stage of development for the type of user service provided, data source acquisition and processing as opposed to data organization, evaluation and dissemination, is performed by all data centers in our study sample. The "young" category of data-oriented information services (many of which might be expected to evolve from the bibliographic-servicing to the data-servicing level) is therefore a potentially important activity.

Texas A&M, the National Bureau of Standards Circular 500, and the JANAF project stated that the provision of references to sources of data for the user is a secondary objective while the remainder of the centers consider provision of references to sources of data one of their primary objectives. Most of the centers provide the user with usable data and information sources: half of the centers also provide the user with usable data without reference to information

sources. In our discussion with the directors and staff of these data compilation activities, they indicated that they devote a substantial portion of their total effort to acquiring and screening sources of data. The magnitudes of these efforts are reflected in the flow charts and internal file descriptions in Appendix C and in the distribution of effort tabulations (presented in Appendix A, pages iii-iv).

Many activities utilize secondary publications, such as abstract services, indices, etc., for locating and acquiring candidate data sources. This often contributes to a substantial time delay between the publication of a new value or bit of data and its availability to the users of the compilation. In some cases data centers have gone to great lengths to formulate screening and acquisition procedures to meet their specific needs. Many times, the information provided the user reflects a compromise between what the compilation director would like to provide and what the economics of the project will permit.

In only a few cases are the content, the format and the recording media of the output of the centers all tailored to the needs of specifically identified users. The X-Ray Diffraction project disseminates their data output in many formats tailored to meet many different users' needs which vary from computer tapes to tabular books and 3 x 5 cards. They also have internal differences in the same output format. The Hanawalt-Davey (tabular) Index lists the three strongest lines, their relative intensities, the chemical formula of the substance, the name of the substance and the accession number of the data card. The Fink Index (tabular) uses the eight strongest lines in order of their relative intensity (cyclically permutated), the name of the substance and the accession number of the data card. Thus this data activitiy can accommodate many different types of users (large corporation, small companies, independent users) and provides a choice of technical presentations of the same data. The ASTM-Wyandotte Spectral Data project is another example of multiple-format output.

C. Type of User Service

All of the activities investigated give as a primary objective the periodic dissemination of accumulated information or data (see Figure 5), while only two activities give as a primary objective the provision of quick response to individual queries. Six chemical data activities identify quick service as a secondary objective or one that is desirable but subordinated by other considerations.

We believe this finding does not imply that the servicing of demand inquiries is a secondary need in chemical data. Rather, probably it reflects limitations in the manpower and monetary resources of the centers, which have settled for programs of periodic publication, guided by an understanding of current needs.

D. Overlap of Coverage

In this study, we found some overlap between centers with respect to substance-property coverage. In the field of physiochemical and thermodynamic properties, the overlap appeared primarily in the basic elements or the hard core data. It seems apparent that this overlap is desirable and convenient from the user's point of view. Many data center directors shared the view that their published documents would be incomplete without this basic data. It would be reasonable to assume that these basic data could be amassed or disseminated by a designated member of the National Chemical Information System. This should decrease significantly the time gap and duplicate effort in evaluation of those numerical property values appearing in the primary literature that find use in the critical compilations of individual scientists and specialized centers.

A typical example of the basic element overlap would be the element oxygen (monoxygen) in the gaseous phase, where JANAF, Texas A & M, and the Contributions to the Data on Theoretical Metallurgy produced by Dr. K. K. Kelley record heat content (enthalpy, heat capacity, entropy and others) for oxygen at 273°-2000°K, while the National Bureau of Standards Circular 500 records heat capacity at a specific temperature within the same range. Using carbon (C) in the gaseous phase, JANAF, Texas A & M, Contributions to the Data on Theoretical Metallurgy, and Selected Values for the Thermodynamic Properties of Metals and Alloys record heat content or enthalpy in similar temperature ranges. Similar comparisons can be shown for other basic elements (nitrogen, hydrogen, sulfur, etc.) or other hard core data.

Examinations of existing compilations indicate that the amount of information stored and/or disseminated varies considerably depending on the nature of the data and of the user populations. Consequently, generalized treatments applied to all chemical substances and all property values should be approached with caution. It is not adequate just to determine which information should be stored; the level (or levels) at which individual data elements can or should be stored needs to be related to usage levels.

E. Nomenclature

Our findings (see Appendix A, pages x-xii) show that all data compilation activities studied index the information under basic substance and property parameters if they have a data source file. Almost all of the systems are keyed to CA nomenclature to the extent that, given the CA name for a substance, the CA coding of that substance could be applied to the system.

Exceptions to this rule occur in the cases of the Cryogenic Data Center, the Thermodynamic Properties of Metals and Alloys Data Center at Berkeley, and

the Atomic and Molecular Processes Information Center. These activities deal with a circumscribed area of substances in a very intensive fashion, and have elaborated either substance description or property description to improve the exactness of retrieval.

Further exceptions appear in the cases of EPIC and TPRC, which deal in part with commercial substances and other mixtures and systems outside the CA Registry nomenclature. These problems, however, are soluble, since each activity has developed structured means for indexing substance coverage of the literature. The work required to make these structures compatible appears reasonable.

Key words for property identification have also been established by most chemical data activities and should prove susceptible to standardization.

A third coordinate used by data compilation activities in indexing data sources for retrieval are measurement environment parameters. The vocabularies and systems are either divergent or non-existent. A good amount of work appears necessary to establish uniform terminology criteria of use to all activities.

F. Data Obsolescence

To date apparently little study has been given to determination of the rate at which specific types of data become obsolete. No centers studied reported established purging elements in their operations procedures. Obviously, this property of data significantly influences the file structure selected to store the data. If chronological age provides a satisfactory criterion for file purging, the process can be easily automated. However, if each bit of data has to be individually examined, the task will require highly sophisticated computer routines, and/or the efforts of a highly skilled data specialist.

G. Purposes of Data Compilation Activities

A common preoccupation of all data compilation activities studies was the desire to improve present expertise and techniques for compressing the voluminous data in the open literature into packages of manageable size for the scientific and technical community. Only shades of difference existed among compilations in their emphases of other purposes. Some of them are mentioned below. In hearings before Congress, it was stated that the Office of Standard Reference Data expected to save $10 to $100 for each dollar expended in data compilation and evaluation efforts. This was estimated to result in a total savings of $100 to $200 million per year. The major portion of the savings was based on a projected reduction of the time scientists and engineers spend in searching for data. Although this is obviously a significant objective of data compilation activities,

it was not the primary justification advanced by center directors in justifying their existence. Most of the more successful compilations are associated with scientific or technical missions that provide the economic justification for the effort. An example is the petroleum industry, which has supported the API Project 44. The government agencies have supported numerous Air Force and other DoD projects, perhaps because firmer estimates of operating savings resulting from the application of more reliable data can be established as offsets against the cost of maintaining the data compilation activity. The association of a data compilation activity with programmatic research or application appears to result in a more consistent funding support, a dynamic service-response posture, and a technical scope tailored to specific identified uses. This suggests that system design concepts which weaken the association between program missions and data compilation activities should be approached with caution.

H. Expertise

Capabilities are being developed by many data compilation activities for generating new data by computations and other projections from the available data base where basic data and knowledge of substance property relationships is sophisticated. The capability requires highly specialized technical people. The need to conserve this competence is a primary concern expressed by key professionals in many of the data compilation activities we interviewed. The capacity to generate data of high quality and broad utility finds important applications in many data compilation activities. For example, the JANAF Thermochemical Data Project stores fundamental thermochemical constants, mathematical formulae and computer routines which are used to generate extensive tables of thermochemical functions. These tables are published and distributed internationally. More advanced processing techniques are being adopted by some activities. JANAF produces punch cards and magnetic tapes containing the fundamental constants with appropriate automated equipment. The ASTM X-Ray Powder Diffraction Data Project stores diffraction data on magnetic tape, and have great processing flexibility for producing customized compilations and indices through manipulation of stored data. They are also experimenting with the generation of the powder patterns by using a computer-plotter setup, or a computer-driven print-out system.

In another area of data generation, the Thermophysical Properties Research Center has developed the expertise to be able to estimate unmeasured property values with a precision adequate to guide design or development efforts until the value could be experimentally determined. The Thermodynamics Research Center utilizes computer techniques for screening evaluated data, and defining and locating their best fit. The resultants are usually represented through im-

provements in the accuracy or precision of the constants of master equations relevant to the property. TRC has also developed computational methods for defining with high accuracy and precision the property-substance curves for all compounds in some homologous hydrocarbon series, based on measurements of reference-purity samples of every fourth compound of the series. Since the preparation of reference-purity samples is an extremely laborious process demanding craft skills few contemporary scientists are willing to develop, such methods may prove critical in the establishment of reference-quality basic data.

## Key Design Considerations for the Data Processing Element of a First-Generation National Chemical Information System

### Decentralized Data Compilation Requirement.

In one of its progress reports to the National Science Foundation, Chemical Abstracts Service posed the question of whether its information system should produce compendia of physical properties information. The first service suggested would provide the name of a substance, the name of the associated physical property, and the literature reference to the source of this information. A proposed later service would include provision of this information, together with associated data necessary for proper interpretation of the property value (e.g., the pressure at which a boiling point was measured.)

Our findings suggest that an alerting service comprising the simpler of the two coverages described by Chemical Abstracts could provide a major relief for one of the largest working burdens found in most of the centers studied. This is the process of searching for relevant documents in the open literature, and their subsequent abstracting, indexing, and coding. A great reduction of essentially duplicated effort in current data-center operations appears feasible by establishing such an operational capability, and coupling it to compilation centers through responsive and expeditious linkages. There is little doubt that centers would welcome such a linkage. Some of them now maintain informal arrangements between themselves to reduce overlapping effort.

The evidence of high scientific and processing expertise in many of the center operations similarly impressed us that subject-specialized data centers possess a sensitivity to their subject-matter that would be most difficult to centralize. They are often centers of scholarly research directed toward advancement of the established structures of the field itself. The data activity they also carry on comprises one of the key tools for this research activity.

We thus suggest that the design thinking for the data component of the National Chemical Information System contemplate a multiplicity of specialized compilation activities suitably linked to other elements of the System.

### Data Source Screening Requirement.

When testifying in support of H. R. 16897, Dr. Donald Hornig, Director, Office of Science and Technology, stated, "The Standard Reference Data System should be considered part of the total national network of scientific and technical information systems." In regard to the question of the interface between the Standard Reference Data System and the activities of the American Chemical Society, Dr. Astin, Director of the National Bureau of Standards, stated, "For example, the first phase of the Standard Reference Data problem is the search for literature. It is also a

problem that is common with the broad scientific information problem. So in improving techniques of search for literature and in making indices of what is in the literature, we (NSRDS and ACS) have a common interest. In this we are working together closely."

As indicated previously, our findings reaffirm a need for improved methods of screening the literature to identify data potentially useful to data compilation activities. We suggest that the system requirement for a national chemical information system include a mechanism for screening data sources and selectively notifying appropriate data compilations when a new or revised property value appears in the literature. Possible means of implementing this system requirement include preparation of substance-property indices and production of citation indices. The publication of citation indices is amenable to automated methods and would provide a valuable tool for the data compiler.

Fast Source Document Processing Requirement.

As noted in our findings, many data compilation activities are currently dissatisfied with the time delays encountered in using secondary sources to locate data sources. A national chemical information processor should acquire source documents and process them in a matter of days rather than months as is the current practice of many indexing services. KWIC indices and title announcements often are not adequate indicators of the data content of articles, reports, etc. Specialized express announcement vehicles appear to be required.

The Data Subsystem.

In the total context of the Chemical Information System operations, we would suggest the following requirements for the data component of the system:

(1) Fast announcement of new or re-measured values for the properties of chemical substances.

(2) Continuous maintenance of a file of the most probable values for a hard-core set of properties of chemical substances.

(3) Effective division of work responsibilities between documentation or information system operators and data specialists.

(4) Maintenance of effective communication between scientists and engineers pursuing mission objectives and the operators of the data compilation activities supporting the pursuit of these specialized missions.

(5) Segregation of data into collections or files of manageable volume so that realistic procedures can be established for screening and purging.

(6) Provision of a workable arrangement for test and evaluation of file integration procedures prior to total integration of a data compilation activity into a broader document-data processing system.

The schematic on the following page illustrates a system configuration which would meet these requirements.

Implementing Requirements.

These suggestions can be implemented through the following development program:

(1) Differentiate work scope of the data component of the National Chemical Information System from the scope of other data systems (e. g., NSRDS, CIDS, and specialized data compilation activities).

(2) Design effective processing linkages between relevant document-processing and data-processing elements of NCIS.

(3) Develop criteria and procedures for rapid and effective announcement of new or re-measured values for the physical and chemical properties of chemical substances.

(4) Develop criteria and procedures for establishing and maintaining an up-to-date file of the most probable values for a "hard-core" set of properties of chemical substances, to provide an effective operational service capability.

(5) Analyze chemical data classes into manageable collections or files, and develop realistic criteria and procedures for data screening, file updating, user servicing, and file purging.

(6) Develop workable criteria and procedures for evaluating specialized document- and data-processing activities that are potential elements of NCIS, and workable policies and procedures for integrating them operationally.

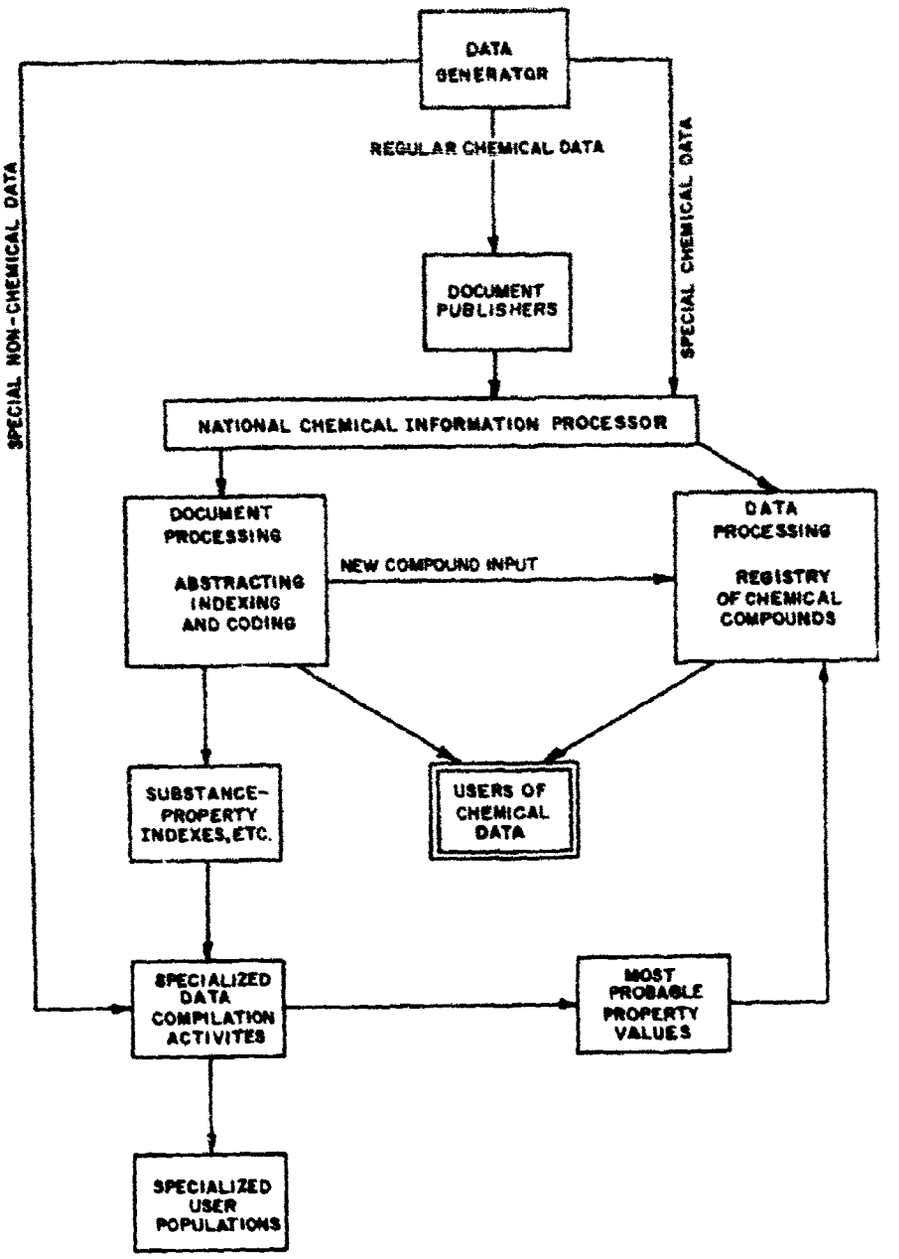
Next Tasks to Undertake.

In its development of the chemical data subsystem, we recommend that the following tasks be undertaken next by the Chemical Information Program:

Task #1: Review the current and projected requirements and capabilities (e. g., current ACS capabilities) for screening sources of data and notifying data compilers and other data users when new or re-measured values for the properties of chemical substances appear in the literature; formulate implementation plans for such services, including where possible, plans for early prototype tests.

FIGURE 6.

Schematic of a Proposed National Chemical Information System
Incorporating Specialized Data Compilation Activities



Task #2: Examine and determine the extent of correlation between:

1. "real-world" file structuring practices as revealed by our data compilation analyses and CIDS user studies;
2. data file contents and structuring as enumerated in the Systems Requirements Description for a National Chemical Information System; and
3. optimum (from a data processing viewpoint) sizes and structurings as determined by National Bureau of Standards Grant GN-455, Goodyear Aerospace Corporation Contract NSF C-456, and University of Pennsylvania Institute for Cooperative Research Contract No. DA18-035-AMC-288, etc.

Task #3: Identify the preferred contents and structuring for the data component of the National Chemical Information System and recommend implementation plans, including plans for prototype tests of file construction and user servicing.

Task #4: Review program description and planning documents (NSRDS, COSATI, etc.) and identify effective interfaces between the data component of the National Chemical Information System and other national scientific and technical data systems.

# APPENDIX A

## Tabulation of Results for the Data Compilation Activities Examined

# TABULATION OF WORKSHEET DATA

## I. OPERATIONAL CONTEXT

### A. Housing Organization

Name

Electrical-Electronic Properties Information Center, Hughes Aircraft Company, Culver City, California (EPIC)

Thermodynamics Research Center, Texas A & M University, College Station, Texas (A & M)

Thermophysical Properties Research Center, Purdue University Research Park, Lafayette, Indiana (TPRC)

Joint Army-Navy-Air Force Thermochemical Tables, Dow Chemical Company, Midland, Michigan (JANAF)

The Wyandotte-ASTM Spectral Data Project, Wyandotte Chemicals Corporation, Wyandotte, Michigan (ASTM)

Thermodynamic Properties of Metals and Alloys Data Center, University of California, Berkeley, California (UC)

Cryogenic Data Center, Cryogenic Engineering Labroatory, National Bureau of Standars, Boulder, Colorado (CDC)

Atomic and Molecular Processes Information Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee (AMPIC)

Chemical Thermodynamic Data Group, Thermochemistry Section, Physical Chemistry Division, Institute of Basic Standards, National Bureau of Standards (NBS)

Joint Committee on Chemical Analysis by Powder Diffraction Methods, American Society for Testing and Materials, Philadelphia (X-RAY)

2. Type of institution:

a. Governmental
NBS
CDC
AMPIC

b. Industrial
JANAF
EPIC

c. Educational
UC
TPRC
A & M

d. Professional Societies
X-RAY
ASTM

3. Primary motivation for data compilation activities:

a. In-house support servicing
TPRC
CDC
AMPIC

b. External servicing, non-profit

| | |
|---|---|
| UC | X-RAY |
| A & M | ASTM |
| JANAF | EPIC |
| TPRC | CDC |
| NBS | AMPIC |

4. Source of funding for data compilation activities:

UC - 10% Industry
90% Government

JANAF - 100% Government

NBS - 100% Government

TPRC - 100% Government

ASTM - 100% Professional Societies from sales of publications

AMPIC - 100% Government

A & M - 13% Industry
35% Government
37% Sales
15% from Texas A & M

CDC - 100% Government

X-RAY - 100% from sales of publications

EPIC - 100% Government

5. Percentage distribution of current data compilation efforts among the following processing functions, e.g.,

UC -
- 5% for acquisition of data sources; for source document screening and evaluation; for source document file (library) maintenance
- 84% for data screening and evaluation; for data extraction
- 10% for compilation publication and dissemination
- 1% for inquiry handling

A & M -
- 5% for acquisition of data sources
- 10% for source document screening and evaluation
- 55% for data screening and evaluation
- 5% for data extraction
- 5% for data file construction and maintenance
- 10% for compilation publication and dissemination
- ~1% for inquiry handling

JANAF-
- 3% for acquisition of data sources
- 5% for source document screening and evaluation
- 2% for source document file maintenance
- 65% for data screening and evaluation
- 5% for data extraction
- 10% for data file construction and maintenance
- 10% for compilation publication and dissemination

NBS -
- 10% for acquisition of data sources; for source document screening and evaluation
- 10% for source document file maintenance
- 40% for data screening and evaluation
- 25% for data extraction; for data file construction and maintenance
- 10% for compilation publication and dessemination
- 5% for inquiry handling

TPRC -
- 32% for acquisition of data sources; for source document screening and evaluation; for source document file maintenance
- 54% for data screening and evaluation; for data extraction; for compilation publication and dissemination
- 14% for inquiry handling

I. A. 5. (cont.)

CDC -
- 35% for acquisition of data sources; for source document screening and evaluation; for source document file maintenance
- 45% for data screening and evaluation; for data extraction; for data file construction and maintenance
- 10% for compilation publication and dissemination
- 10% for inquiry handling
- >1% for other processing functions

X-RAY -
- 3% for acquisition of data sources
- 6% for source document screening and evaluation
- 30% for data screening and evaluation
- 7% for data extraction
- 40% for data file construction and maintenance
- 10% for compilation publication and dissemination
- 4% for other processing functions

ASTM -
- 90% for data file construction and maintenance; for compilation publication and dissemination
- <1% for inquiry handling

EPIC -
- 33% for acquisition of data sources; for source document screening and evaluation; for source document file maintenance
- 33% for data screening and evaluation; for data extraction
- 33% for compilation publication and dissemination; for inquiry handling; for other processing functions

AMPIC -
- 60% for acquisition of data sources; for source document screening and evaluation; for source document file maintenance
- 30% for data screening and evaluation; for data extraction; for data file construction and maintenance
- 5% for compilation publication and dissemination
- 5% for inquiry handling

## II. DATA SOURCE (LIBRARY) COLLECTION AND PROCESSING

### A. Identification and Acquisition of Sources of Data

1. Which of the following are screened in searching for candidate document titles or data sources? *

a. Journal Articles:

| | | |
|---|---|---|
| UC - | (Approx.) | 5% |
| A & M - | | 15% |
| JANAF - | | 5% |
| NBS - | | 60% |
| TPRC - | | 65% |
| CDC - | | 45% |
| X-RAY - | | 100% |
| EPIC - | | 65% |
| AMPIC - | | 90% |

b. Catalog Cards:

| | | |
|---|---|---|
| CDC - | (Approx.) | 2% |

c. Bibliographic Citations in Indexed documents:

| | | |
|---|---|---|
| UC - | (Approx.) | 2% |
| A & M - | | 20% |
| NBS - | | 2-3% |
| TPRC - | | 24% |
| CDC - | | 15% |

d. Index and Abstract Publications:

| | | |
|---|---|---|
| UC - | (Approx.) | 25% |
| A & M - | | 55% |
| JANAF - | | 5% |
| NBS - | | 35% |
| TPRC - | | 30% |
| CDC - | | 20% |
| EPIC - | | 25% |
| AMPIC - | | 10% |

e. Computerized Tape Indexes:

| | |
|---|---|
| JANAF - | 85% |

f. Punched Card Indexes:

None

g. Others:

| | | |
|---|---|---|
| UC - | (Approx.) | 8% |
| A & M - | | 10% |
| JANAF - | | 10% |
| NBS - | | 2% |
| TPRC - | | 18% (for current contents) |
| EPIC - | | 5% |

* X-RAY and ASTM do not have data source files.

2. Total number of candidate sources identified annually:

| | |
|---|---|
| UC - | 1400 |
| JANAF - | 1500 |
| NBS - | 2500 |
| TPRC - | ~9200 |
| CDC - | 9100 |
| EPIC - | 212000 |
| AMPIC - | 3500 |

3. Which of the following data sources are annually indexed?

a. Journal Articles:

| | | |
|---|---|---|
| UC - | (Approx.) | 80% |
| A & M - | | 80-85% |
| JANAF - | | 90% |
| NBS - | | 85% |
| TPRC - | | 76% |
| CDC - | | 60% |
| EPIC - | | 76.8% |
| AMPIC - | | 75% |

b. Published Reports other than Journal Articles:

| | | |
|---|---|---|
| UC - | (Approx.) | 10% |
| A & M - | | 5% |
| JANAF - | | 10% |
| NBS - | | 12% |
| TPRC - | | 24% |
| CDC - | | 25% |
| EPIC - | | 12.6% |
| AMPIC - | | 25% |

c. Unpublished Laboratory Reports:

| | | |
|---|---|---|
| UC - | (Approx.) | 6% |
| A & M - | | 5% |
| CDC - | | 13% |

d. Data Sheets from other Compilations:

| | | |
|---|---|---|
| UC - | (Approx.) | 4% |
| A & M - | | 5% |
| CDC - | | 2% |
| EPIC - | | 7.4% |

e. Customized Data Sheets:

| | | |
|---|---|---|
| NBS - | (Approx.) | 3% |

f. Others:

| | | |
|---|---|---|
| EPIC - | (Approx.) | 3.2% |

4. What is the form of data sources currently accessioned?

a. Hardcopy:

| | | |
|---|---|---|
| UC - | (Approx.) | 75% |
| A & M - | | 100% |
| JANAF - | | 100% |
| NBS - | | 90% |
| TPRC - | | 80% |
| CDC - | | 70% |
| X-RAY - | | 100% |
| ASTM - | | 100% |
| EPIC - | | 100% |
| AMPIC - | | 90% |

b. Microcopy:

| | | |
|---|---|---|
| UC - | (Approx.) | 25% |
| NBS - | | 10% |
| TPRC - | | 20% |
| CDC - | | 30% |
| AMPIC - | | 10% |

5. What is the total number of data sources indexed to date?

| | |
|---|---|
| UC - | 12,000 |
| A & M - | 30,000 |
| JANAF - | 26,000 |
| NBS - | 55,000 |
| TPRC - | 33,700 |
| CDC - | 40,000 |
| EPIC - | 27,000 |
| AMPIC - | 6-7,000 |
| X-RAY - | 13,000 |

6. What is the period covered by data sources indexed to date?

| | |
|---|---|
| UC - | from 1848 |
| A & M - | 1840 |
| JANAF - | 1925 |
| NBS - | 1940 (expanded in) |
| TPRC - | 1920-1964 |
| CDC - | from 1950 |
| X-RAY - | 1940 |
| ASTM - | unknown |
| EPIC - | unknown |
| AMPIC - | from 1950 |

7. What percentage of the total, usable data sources issued during the period of coverage has been indexed?

| | |
|---|---|
| UC - | 85% |
| A & M - | 100% |
| JANAF | >90% |
| NBS - | 97% |
| TPRC - | 99% |
| CDC - | 60% |
| EPIC - | unknown |
| AMPIC - | 95% |

8. What percentage of currently issued, usable data sources is indexed?

| | |
|---|---|
| UC - | 95% |
| A & M - | 100% |
| JANAF - | >90% |
| NBS - | 90% |
| TPRC - | 99% |
| CDC - | 85% |
| EPIC - | 90% |
| AMPIC - | 100% |

9. What percentage of current year accessions were issued or published within the past year?

| | |
|---|---|
| UC - | ~1% |
| A & M - | 50% |
| JANAF - | 90% |
| NBS - | 87% |
| TPRC - | 50% |
| CDC - | 85% |
| EPIC - | unknown |
| AMPIC - | 85-90% |

## B. Data Source Screening & Evaluation

1. What are principal criteria for screening of titles?

| | |
|---|---|
| UC - | Thermodynamic articles on metals and alloys and phase diagram data. |
| A & M | Judgment by a specialist, Chemical Abstract per entry indicates relevant scope. |
| JANAF - | Measurement of property of interest to thermal lab.; screening by weighted key words. |
| NBS - | Search all organics; publish only those with two or fewer Carbon Atoms; extraction from original documents. |
| TPRC - | Checking for property coverage and duplication. |
| CDC - | Primary temperature ranges below -200° F. |
| EPIC - | Experimental data on the E. P. I. C. |
| AMPIC - | Atomic and molecular physics |

2. What percentage of initially considered titles is rejected without physical examination of the data source?

| | |
|---|---|
| UC - | 75% |
| A & M - | None |
| JANAF - | >95% |
| NBS - | 8% |
| TPRC - | NA |
| CDC - | 40% |
| EPIC - | 90% |
| AMPIC - | None |

3. What are the principal criteria for screening of data sources?

| | |
|---|---|
| UC - | Have to have Thermodynamic or equilibrium information. |
| A & M - | Pure samples, sound experimental measurements, reputation of author. |
| JANAF - | New measurements sought, remeasurement of property value is accepted. |
| NBS - | Pertinence - Chemical thermodynamic data on pure substances and mild electrolytic solutions. |
| TPRC - | Pertinence, duplication. |
| CDC - | Temperature range - 40% is screened out because it is probably non-technical. |
| EPIC - | Duplication; whether actual data is there, source quality, suitable subject matter. |
| AMPIC - | Retained if contains useful data. |

4. What percentage of data sources physically examined is rejected?

| | |
|---|---|
| UC - | 2% |
| A & M - | None |
| JANAF - | 50% |
| NBS - | 2-3% |
| TPRC - | 50% |
| CDC - | 28% |
| EPIC - | 75% |
| AMPIC - | 10-20% |

C. Data Source Cataloging and Indexing

1. Indicate the extent to which the modes of cataloging and indexing are used.

UC - a. 90 substance and 23 properties and type of measurement descriptors used to date; b. none added annually; c. average of 6 substance and 5 properties and type of measurement descriptors assigned to each document.

A&M - a. 20,000 substance indication and 60-70 property identification descriptors are used to date, 140-150 types of measurement descriptions are given, but the system is not organized on this parameter and they have not been reduced to "terms"; b. approximately 200 substance, 2-3 property, and 5 types of measurement are added annually; c. 5-10 substance, 3-5 property, and 3-5 type of measurement descriptors are assigned to each document; d. approximately 3,000 file entries per year.

JANAF - a. 10,000 substance identification, 70 property identification, 18 type and methods of measurement, and 9 substance characterization descriptors are used to date; b. approximately 1,000 substance identification descriptors are added per year; c. an average of 5 descriptors are assigned to each source document; d. 7,500 descriptor entries annually.

NBS - Approximately 40,000 substance assignments used to date; properties grouped into 10 categories, with 13 subcategories; type of measurement, temperature range etc., not reduced to "terms" but is specified in content; 10,000 file entries per year.

TPRC - a. 45,116 substance, 7 property, 7 treatment of subject, 9 substance state, and 9 language of article descriptors are used to date; b. rate of addition of substance descriptors is not available, property, subject treatment, substance state, and language descriptors are fixed in number; c. an average of 2.5 to 3 substance and property descriptors assigned to each source document, 1 subject, 1 substance state, and 1 language descriptor; d. approximately 12,650.

II. C. 1. (cont.)

| | |
|---|---|
| CDC - | a. 500 substance and 350 property descriptors (includes variations of more basic properties), 550 conditions; b. 50 substance, 10 property, and 50 condition descriptors added annually; c. 3 to 4 substance, 1 to 2 property and 4 to 6 condition descriptors assigned to each source document. |
| X-RAY - | N/A |
| ASTM - | N/A |
| EPIC - | a. 5100 substance identification and 56 property identification descriptors used to date; b. 350 substance descriptors added annually, properties added infrequently; c. average number of descriptors for each document not available; d. descriptor entries not available. |
| AMPIC - | a. unlimited number of substance identification descriptors used to date, no property identifications used, 99 Process reactants used to date; b. unlimited substance ident. in current annual rate of addition of new descriptors but no other means of adding descriptors given; c. unlimited average number of descriptors assigned to each source document in substance identification, no other descriptors assigned; d. unlimited growth of descriptor entries for substance ident., no other growths shown. |

2. Indicate which means are used to index the substance coverage of data sources:

| | |
|---|---|
| A&M - | C. A. or other systematic chemical nomenclature and OCT Standard Order of Arrangement. |
| UC - | Chemical fragment codes. |
| JANAF - | C. A. or other systematic chemical nomenclature. |
| NBS - | C. A. or other systematic chemical nomenclature, standard order of arrangement by empirical formula. |
| TPRC - | C. A. or other systematic chemical nomenclature, in combination with organization rules based on chemical composition and classes of substances. |
| CDC - | Trivial name of compounds, compleses, systems, etc. |
| EPIC - | C. A. or other systematic chemical nomenclature. |
| AMPIC - | C. A. or other systematic chemical nomenclature. |
| X-RAY- | N/A |
| ASTM - | N/A |

3. What nomenclature standards, thesauri, code manuals, etc. are used to control indexing of substance coverage?

| | |
|---|---|
| UC - | index cards, controlled vocabulary using. |
| A&M - | IUPAC and CA Recommendations. |
| JANAF - | modified Hill Indexing System. |
| NBS - | IUPAC, Chemical Abstracts. |
| TPRC - | Dictionary of Substances - numerical code designations. |
| CDC - | Characteristic Coding Guide. |
| X-RAY - | N/A |
| ASTM - | N/A |
| EPIC - | from Trade Name to Chemical Name. |
| AMPIC - | no control; Index all molecules of 5 atoms and under. |

4. What means are used to index the property coverage of data sources?

| | |
|---|---|
| UC - | indexing card, controlled vocabulary. |
| A&M - | as originally standardized by Rossini. |
| AMPIC - | a standardized thesaurus. |
| NBS - | list of properties - 10 catagories. |
| TPRC - | list of 7 properties (actually covers 13) |
| CDC - | own terminology with hundred relationships. |
| X-RAY - | N/A |
| ASTM - | N/A |
| EPIC - | Indexing Guide or Glossary of Electronic Properties (56 terms and 400 Use References) |
| JANAF - | 70 property identification key words, 18 key words indicating measurement, and 9 key words indicating the state or condition of the substance. |

D. Physical Arrangement of Data Source Files (Library) See Appendix C

E. Purging of Data Source File (Library):

1. Indicate the method of purging data source file (library):

| | |
|---|---|
| UC - | file is not purged. |
| A&M - | no data source file. |
| JANAF - | file is not purged. |
| NBS - | file is not purged. |
| TPRC - | file is not purged. |
| CDC - | file is not purged. |
| X-RAY - | N/A |
| ASTM - | N/A |
| EPIC - | file is not purged except for errors or duplication. |
| AMPIC - | file is not purged. |

F. If the Data Source File Supports More Than One Data File or Compilation, Identify the File and Compilation (s) See Appendix B

G. Flow Chart the Steps Involved in Data Source File Construction and Maintenance See Appendix B

**Science Communication**
Washington, D. C. 200 07



## III. DATA FILE CONSTRUCTION

### A. Screening and Evaluation

1. What criteria are used to select data for inclusion in the data file?

   a. None, except that the data fall within the scope of the data file and are from an apparently reliable source.

      A & M
      JANAF
      NBS
      ASTM

   b. Data which meets the file's criteria for substance purity, measurement reliability, etc.

      X-RAY

   c. Critically qualified data which meet the "one best estimate" value for the property.

      CDC
      X-RAY

   d. Not applicable.

      UC
      TPRC
      EPIC
      AMPIC

2. When is data screened for exclusion from the data file?

   a. Prior to extraction from the source document.

      JANAF
      ASTM

   b. Subsequent to extraction from the source document.

      CDC
      X-RAY

   C. Not applicable.

      UC
      A & M
      NBS
      TPRC
      EPIC
      AMPIC

A. (cont'd.)

3. What percentage of source documents screened provide data useful in data file construction?

JANAF - >90%
NBS - 95%
CDC - 25%
X-RAY - 65%
ASTM - >95%

B. File Accumulation and Purging

1. What is the earliest published date covered by data items processed to present?

A & M - 1840
JANAF - 1925
NBS - 1780
CDC - 1960
X-RAY - 1940
ASTM - 1949

2. What is the total number of data sources processed to date, i.e., what is the total number of documents or data sources from which data have been extracted?

A & M - 27,000
JANAF - 25,000
NBS - 55,000
X-RAY - 13,500
ASTM (R) - 95,000
ASTM (V) - 25,000
ASTM (mass spec.) - 3,300

3. What percentage of potentially useful data sources (see item II-A-6) has been screened for processing of data into the file?

A & M - 10%
JANAF - 100%
NBS - ~80%
CDC - >95%
X-RAY - 100%
ASTM - 100%

B. (cont'd.)

4. What combinations of substances and properties are normally found in source documents?

   a. Single Substance:

   A & M - 20%
   JANAF - 70%
   NBS - 30%
   X-RAY - 70%
   ASTM - 80%

   b. Multiple Substances:

   A & M - 80%
   JANAF - 30%
   NBS - 70%
   X-RAY - 30%
   ASTM - 20%

   c. Single Property:

   A & M - 40%
   JANAF - 90%
   NBS - 60%
   X-RAY - -
   ASTM - 100%

   d. Multiple Properties:

   A & M - 60%
   JANAF - 10%
   NBS - 40%
   X-RAY - 100%
   ASTM - -

5. What is the current annual rate of processing data into the file?

   A & M: 700 additional source items included; approximately 200 additional substances included; 2-3 additional properties; 500 new values and/or properties not previously included; 2,500 additional values for substances and properties previously included.

   JANAF: 1,500 additional source items included; approximately 1,000 additional substances included; no additional properties included; number of additional values not available.

B. (cont'd.)

NBS: 2,200-2,500 additional source items included; 1,000 additional substances included; no additional properties included; 1,500 cards of new values; 8,500 cards of additional or updated values for substances included.

X-RAY: 1,200 additional substances covered.

ASTM: 10,000 Infrared spectra per year; 4,000 Ultraviolet and visible spectra per year.

6. What percentage of data extracted is in each of the following forms:

a. Digital or alphanumeric

A & M - 100%
JANAF - 85%
NBS - 70%
TPRC - 20%

b. Graphic

JANAF - 10%
TPRC - 80%
X-RAY - 100%
ASTM - 100%

c. Formula

JANAF - 5%
NBS - 15%

d. Other

NBS - 15%

7. What is the frequency of updating of the data file?

a. Continuously

A & M
JANAF
NBS
CDC
X-RAY
ASTM

b. N/A

TPRC

B. (cont'd.)

8. How are data files purged?

a. Files are not purged.

NBS

b. Property values are purged when more reliable values become available.

CDC
X-RAY
ASTM

c. Files are retired when a data compilation is published.

JANAF

d. Not applicable.

A & M
TPRC

C. Chemical Substance Coverage

1. What is the intended substance coverage for this file?

A & M: API pure hydrocarbons and related compounds;
TRC(MCA) - primarily organic and inorganic compounds outside API and of interest fo the chemical industry

JANAF: All chemical compounds of interest to Dow Chemical Company

NBS: All inorganic (pure) substances and all pure organic substances and well-defined mixtures thereof.

TPRC: All substances

CDC: Cryogenic fluids (light hydrocarbons, atmospheric gases, rare gases)

X-RAY: Inorganic and organic which will produce diffraction patterns; inorganic = metals, alloys, minerals.

ASTM: Infrared, Ultraviolet spectroscopy and Mass Spectrometry; Mostly organic compounds

C. (cont'd.)

2. What is the total number, to date, of chemical substances included in this data file?

| | |
|---|---|
| A & M: | 20,000 |
| JANAF: | 10,000 |
| NBS: | 40,000 |
| TPRC: | 42,900 |
| X-RAY: | 13,500 |
| ASTM- | |
| IR: | 95,000 |
| UV: | 25,000 |
| Mass spec.: | 3,300 |

3. What is the distribution of coverage among the following groups of chemical substances?

A & M: 2% single chemical elements; 10% inorganic compounds; 87% organic compounds; less than 1% organic mixtures and compounds.

JANAF: Less than 5% chemical elements; 47.5% inorganic compounds, mixtures and systems; 47.5% organic compounds, mixtures and systems.

NBS: 4% chemical elements; 3% single chemical elements; 1% elemental mixtures and systems; 40% inorganic compounds, mixtures and systems; 25% inorganic compounds; 15% inorganic mixtures and systems; 56% organic compounds, mixtures and systems; 50% organic compounds; 6% organic mixtures and systems.

TPRC: N/A

X-RAY: 26% chemical elements (2% single chemical elements, 98% elemental mixtures and systems); 37% inorganic compounds, mixtures and systems (70% inorganic compounds, 30% inorganic mixtures and systems); 37% organic compounds, mixtures and systems (90% organic compounds, 10% organic mixtures and systems).

ASTM: 10% inorganic compounds, mixtures and systems; 90% organic compounds, mixtures and systems.

D. Property Coverage

1. What is the intended property coverage for this file?

   A & M: Simple physical properties, thermochemical data, ideal gas thermal data, viscosity.

   JANAF: Thermochemical and related fundamental properties.

   NBS: Chemical thermodynamic properties.

   TPRC: N/A

   CDC: Thermodynamic transport (PVT) relationships, specific heat and entropy, dielectric constant.

   X-RAY: Interplanar spacings, relative intensities, radiation source, wave length, filter, crystallographer system, space group, lattice parameters, interaxial angles, chemical formula, number of atoms per unit structure, indices of refraction, measured density, melting point, color, hardness, luster, chemical analysis, source, heat treatment, temperature, name.

   ASTM: Infrared and Ultraviolet spectral data; mass spectral data.

2. What is the total number, to date, of properties included in this data file?

   A & M: 60-70
   JANAF: 70
   NBS: 95
   TPRC: -
   CDC: 15
   X-RAY: 20-25
   ASTM-
   IR: 95,000
   UV: 25,000
   Mass spec: 3,300

3. How many new properties are annually added to the file?

   A & M: 2-3
   JANAF: 0
   NBS: 0
   TPRC: -
   CDC: 2
   X-RAY: < 1%
   ASTM-
   IR: 30-40
   UV: 15-25
   Mass spec: 15-25

D. (cont'd.)

4. What is the average number of values per property-substance included in the file?

   A & M: 25

   JANAF: N/A

   NBS: 25

   TPRC: N/A

   X-RAY: 15-20

   ASTM: 30-40

F. File Formatting

1. Is the source document format retained or is the format converted to a standardized format?

   a. Format retained:

      JANAF (90%)

   b. Format converted:

      A & M
      JANAF (10%)
      NBS
      CDC
      X-RAY
      ASTM

   c. Not applicable:

      TPRC

2. What is the current volume of the file?

   a. File units:

      A & M: 200,000
      JANAF: 8,000
      NBS: ~200,000

   b. Data records:

      X-RAY: in 16th supplement

   c. Not applicable:

      TPRC
      CDC
      ASTM

## IV. USER SERVICING

### A. Service Philosophies and Emphases

1. How are the following service features rated in your current operations? (see figure 5)

2. Who are the users of the products of this data compilation activity?

| | Basic Scientists | Applied Technologists | |
|---|---|---|---|
| UC | 40% | 60% | |
| A & M | 50% | 50% | |
| JANAF | 25% | 75% | |
| NBS | 40% | 60% | |
| TPRC | 15% | 80% | 5% (mgmt) |
| CDC | 20% | 80% | |
| X-Ray | 20% | 80% | |
| ASTM | 30% | 70% | |
| EPIC | 50% | 50% | |
| AMPIC | 75% | 25% | |

### B. Types of Output Prepared

1. Indicate the distribution of volume of output:

| | Data source reference without data extracts | Data extracts without data source references | Data extracts and data source references |
|---|---|---|---|
| UC | | | 100% |
| A & M | <1% | | 99% |
| JANAF | | | 100% |
| NBS | 5% | | 95% |
| TPRC | | NA | |
| CDC | 30% | | 70% |
| X - Ray | | | 90 - 100% |
| ASTM | 80% | | 20% |
| EPIC | 85% | | 15% |
| AMPIC | 100% | | |

2. Indicate the distribution of output according to the type of requirement:

| | Generated in Response to Inquiries | | Generated for Periodic Publication | |
|---|---|---|---|---|
| UC | 1% | | 99% | |
| A & M | | | 100% | |
| JANAF | | | 100% | |
| NBS | 2% | | 98% | |
| TPRC | | NA | | |
| CDC | 10% | | 85% | 5% (other) |
| X - Ray | | | >99% | |
| ASTM | | | 99% | |
| EPIC | ~60% | | ~40% | |
| AMPIC | | | 100% | |

C. Usage of Data Source Files (Library)

1. Indicate the frequency of usage of the data source file (library).

UC - 10 searches per year in response to external queries for for document references or to identify sources of data required to answer data queries; daily searches by staff to identify sources of data required to prepare or update data compilations.

A & M -100 searches in response to external queries for document references and/or data; continual searches by staff to identify sources of data required to prepare or update data compilations (the same file is used for IV C 1. and IV D).

JANAF -4 magnetic tape searches and 100 manual searches of file folders are made by staff to identify sources of data required to prepare or update data compilations.

NBS - 20 searches per year in response to external queries for document references, continuous searches by staff to identify sources of data required to prepare or update data compilations.

C. (con't)

TPRC - Daily manual searches by staff in response to external queries for document references or to identify sources of data required to answer data queries; periodic searches every 12 to 18 months for preparation or updating of a bibliography or accession list for external distribution (Retrieval Guide); continuous searches by staff to identify sources of data required to prepare or update data compilations.

CDC - Approximately 90 searches in response to external queries extensive bibliography of references, approximately 200 searches by staff to identify sources of data required to answer specific data queries, approximately 10 to 15 searches by staff to identify sources of data required to prepare or update data compilations.

X-Ray - not applicable

ASTM - not applicable

EPIC - 75 - 80 searches per year in response to external queries for documents references; 1 search per year for preparation or updating of a bibliography or accession list for external distribution; 7 or 8 searches by staff to identify sources of data required to prepare or update data compilation.

AMPIC - 30 searches per month in response to external queries for document references; 5 searches per year for periodic searches for preparation or updating of a bibliography or accession list for external distribution; no searches by staff to identify sources, no searches by staff to identify sources of data.

2. List accession lists or bibliographies prepared and disseminated, frequency of publication, forms distributed and type of Indexing.

| | Accession List or Bibliography | Frequency of Publication or updating | Forms Distributed | Type of Indexing |
|---|---|---|---|---|
| UC - | N/A | – | – | – |
| A&M | IuPAC Bibliography | annual | hardcopy | substance-property index & guide |
| JANAF - | N/A | – | – | – |
| NBS - | IuPAC Bibliography | yearly | hardcopy | substance-property index & guide |
| CDC - | Current Awareness Service | weekly | hardcopy | broad subject; catalogue |
| | Bibliographies | irregularily | hardcopy | all on one subj. (by product of compilation efforts) |
| | or | on request | hardcopy | on topics requested |
| | Accession Lists | monthly | hardcopy | none |
| X-RAY - | N/A | – | – | – |
| ASTM - | N/A | – | – | – |
| EPIC - | Interim Reports | in response to inquiries | hardcopy | substance-property |
| | Bibliographies | in response to inquiries | hardcopy | substance-property |
| AMPIC - | Bibliographies | semi-annual | hardcopy | reactants and processes |
| TPRC - | Retrieval Guide to Thermophysical Properties Research Literature | 12-18 months | Bound Volumes | substance-property, type of subject treatment |
| | Bibliographies | on request | Computer Printout | substance-property specified |

4. Is the format of the data source file (library) normally retained or modified for bibliography?

| | |
|---|---|
| UC - | N/A |
| A&M - | separate bibliography cards - no data source file per se |
| JANAF - | N/A |
| NBS - | modified |
| TPRC - | retained |
| CDC - | retained |
| X-RAY - | N/A |
| ASTM - | N/A |
| EPIC - | retained |
| AMPIC - | retained |

D. Usage of Data File

| | | |
|---|---|---|
| UC - | UC - | N/A |
| | A&M - | 100 searches in response to external queries for document references and/or data; continual searches by staff to identify sources of data required to prepare or update data compilations. |
| | JANAF - | Data file not used for searches; it is part of a compilation scheme. |
| | NBS - | 200 searches per year to obtain data for data queries, continuous searches to obtain data to prepare or update compilations. |
| | CDC - | Approximately 35 searches to obtain data for data queries, 10 searches to obtain data to prepare or update compilations. |
| | TPRC - | N/A |
| | X-Ray - | Very few searches to obtain data for data queries; searches to obtain data required to prepare or update data compilation would only take place when converting cards to Book Form and when replacing an old pattern with a new one. |
| | ASTM - | Once a year for each tabular publication; constant merging. |
| | EPIC - | N/A |
| | AMPIC - | N/A |

E. Publication Statistics (See Appendix B)

F. Screening and Evaluation for Compilations

1. What criteria are used to select data for inclusion in the Compilation(s)?

UC - Data which meet the compilation's criteria for substance purity, measurement reliability, etc.; critically qualified "one best value", (refers to Selected Values of Thermodynamic Properties of Metals & Alloys, and Supplements)

A&M - Critically qualified "one best value"(refers to API Selected Values of Properties of Hydrocarbons and Related Compounds, and TRC Selected Values of Properties of Chemical Compounds)

Data which meet the compilation's criteria for substance purity, measurement reliability, etc; maximum of 3 spectra are included for each substance (refers to API and TRC Selected Spectral Data Catalogues)

JANAF - Critically qualified "one best Value" (refers to JANAF Thermochemical Tables)

NBS - Critically qualified "one best value" (refers to Selected Values of Chemical Thermodynamic Properties and NBS Technical Notes)

TPRC - None, except that the data fall within the coverage of the data compilation and are from a reliable source; critically qualified "recommended values" (refers to TPRC Data Book)

X-RAY - Data which meet the compilation's criteria for substance purity, measurement reliability, etc. and some critically qualified "one best value" data (see Appendix B for reference)

ASTM - None, except that the data fall within the coverage of the data compilation and are from an apparently reliable source (refers to Infrared, Ultraviolet and Mass Spectral Data Indexes)

EPIC - Data which meet the compilations criteria for substance purity, measurement reliability, etc.; primary sources only (refers to Monographs)

AMPIC - N/A

CDC - Data which meet the compilations criteria for substance purity, measurement reliability, etc. for preliminary compilations; Critically qualified "one best value" (refers to Monographs)

2. What percentage of data pulled from the data file is normally accepted for inclusion in the data compilation?

| | |
|---|---|
| UC - | N/A |
| A&M - | Approximately 25% (refers to API Selected Values and TRC Selected Values) |
| | N/A (refers to API and TRC Selected Spectral Data Catalogues) |
| JANAF - | 100% |
| NBS - | Not available |
| TPRC - | N/A |
| X-RAY - | 100% |
| ASTM - | 100% |
| EPIC - | N/A |
| AMPIC - | N/A |
| CDC - | N/A |

3. What percentage of potentially useful data items in the data file have been considered for inclusion in the data compilation?

| | |
|---|---|
| UC - | 100% |
| A&M - | 100% (API - Selected Values and TRC - Selected Values) |
| | N/A (API and TRC Selected Spectral Data Catalogues) |
| JANAF - | >95% |
| NBS - | 90% |
| TPRC - | 100% |
| X-RAY - | Not available |
| ASTM - | Not available |
| EPIC - | 100% |
| AMPIC - | Not available |
| CDC - | Not available |

4. What is the current annual rate of issuing data in compilations?

UC - Some new substances included each year, some revised or updated annually.

A&M - N/A (for all publications)

JANAF - 150 additional source items included, 100 additional substances included, no additional properties included, values updated for 100 substances.

NBS - Complete revision is in process.

TPRC - 1, 391 additional source items included, 3, 725 additional specimens included, no additional properties, 1086 pages on new substances and/or properties included, and 642 additional or updated pages for substances and properties previously covered.

X-RAY- Same as data file - 1, 200 additional substances covered.

ASTM - Not available

EPIC - N/A

AMPIC - N/A

CDC - N/A

G. Chemical Substance Coverage of Compilations

1. What is the intended substance coverage of this compilation?

UC - metals and alloys.

A&M - Pure hydrocarbons and related compounds (such as oxygen and sulfur derivatives of hydrocarbons) (refers to API Selected Values)

Pure organic and inorganic substances outside of the hydrocarbon (API) area (refers to TRC Selected Values)

API Spectra cover Hydrocarbons and related compounds MCA spectra cover other organic and inorganic substances - coverage is not complete (refers to API and TRC Selected Spectra Data)

JANAF - light elements and their oxides, fluorides, etc.

NBS - all inorganic and organic substances with molecules of 2 or less atoms.

TPRC - all substances

X-RAY - Not available

EPIC - super conductors, semiconductors, insulators, metals, ferroelectrics, ferromagnetics, ferrites, electroluminescents, thermo emitters.

ASTM - Not available

AMPIC - Not available

CDC - Cryogenic Substances.

2. What is the total number, to date, of chemical substances included in this data compilation?

UC - Approximately 75 elements, 300-400 alloys

A&M - 3375 (API - Selected Values)

2000 (TRC - Selected Values)

not available (API-TRC Selected Spectral Data)

JANAF - 625 (estimated)

NBS - 4,500 chemical compounds

TPRC - 7,700 specimens, substances not applicable.

X-RAY - 13500 (8500 inorganic, 5000 organic)

ASTM - IR-95,000; UV-25,000; Mass Spec-3,300

EPIC - Approximately 50

AMPIC - N/A

CDC - 20 substances (principal fluids and common mixtures of these fluids) plus metallic elements and selected alloys.

3. What is the distribution of coverage among the following groups of chemical substances?

UC - 20% single chemical elements, 80% elemental mixtures and systems.

A&M - 100% inorganic compounds (API-Selected Values)

<2% chemical elements, 12% inorganic compounds, 85% organic compounds (TRC-Selected Values)

Almost 100% organic, few inorganic (API-TRC SELECTED SPECTRAL)

JANAF - 10% Chemical elements, 90% inorganic compounds, mixtures &syst.

NBS - Approx. 3% chemical elements, 87% inorganic, 10% organic compounds

TPRC - Not available

X-RAY - Not available

ASTM - 10% inorganic compounds, mixtures and systems, 90% organic compounds, mixtures and systems.

EPIC - Covers all chemical substances concerned with electrical-electronic properties.

AMPIC - Covers all chemical substances concerned with molecular and atomic processes

CDC - 100% chemical elements, mixtures and systems.

H. Property Coverage of Compilations

1. What is the intended property coverage for this compilation?

UC - Thermodynamics, properties, phase equilibria

A&M - Thermodynamic properties (API-Selected Values)

Physical and thermodynamic properties (TRC-Selected Values)

Spectral properties (API-TRC Selected Spectral Data)

JANAF - Heat, capacity, entropy, free energy function, enthalpy, heat of formation, equilibrium constant

NBS - Enthalpy of formation, Gibbs Energy of formation, Entropy, Heat capacity

TPRC - Conductivity, viscosity, thermal diffusivity, specific heat

X-RAY - Not available

ASTM - Not available

EPIC - Electronic Properties

AMPIC - Not available

CDC - Not available

2. What is the total number, to date, of properties included in this compilations?

UC - 20 distinct properties, not including properties redundantly expressed

A&M - Approx. 20 (API and TRC Selected Values)

1 each (API-TRC Selected Spectral Data)

JANAF - 2

NBS - 12

TPRC - 5

X-RAY - 20 to 25, approximately for each publication

ASTM - UV, and Mass Spectra = 3

EPIC - 56

AMPIC - N/A

CDC - 15 for thermodynamic properties of cryogenic fluids.

3. What is the average number of values per property-substance included in the compilation?

| | |
|---|---|
| UC - | Elements - 11 alloys, increases with different temperature and composition intervals. |
| A&M - | 15 (for each API & TRC Selected Values) |
| | 1 to 3 (API-TRC Selected Spectral Data) |
| JANAF - | 61 |
| NBS - | 1 |
| TPRC - | N/A |
| X-RAY - | 15 to 20 for each publication |
| EPIC - | N/A |
| ASTM - | N/A |
| AMPIC - | N/A |
| CDC - | N/A |

I. Compilation Formatting

1. Is the data file format retained or is the format converted to a different publication or dissemination format?

| | |
|---|---|
| UC - | format converted |
| A&M - | format modified |
| JANAF - | format retained |
| NBS - | format converted |
| TPRC - | format retained |
| X-RAY - | not available |
| ASTM - | not available |
| EPIC - | format retained |
| AMPIC- | not available |
| CDC - | not available |

2. What is the current volume of the data compilation?

| | |
|---|---|
| UC - | N/A |
| A&M - | 2391 data sheets, 400,000 information units (API Selected Values) |
| | 780 data sheets, >28,000 information units (TRC Selected Values) |
| | Infrared Spectral Data, TRC-302 and API-2650 data sheets; Ultraviolet Spectral Data, TRC-103 and API-1014 sheets; Raman Spectral Data, TRC-32 and API-456 sheets; MassSpectral Data, TRC-149 and API-2191 sheets; Nuclear Magnetic Resonance Spectral Data, TRC-492 and API-583 sheets (API-TRC Selected Spectral Data) |
| JANAF - | 950 data sheets |
| NBS - | 1268 pages. |
| TPRC - | 3,322 data sheets |
| X-RAY - | 8,500 compounds in each inorganic compilation<br>5,000 compounds in each organic compilation |
| ASTM - | Approximately 123,300 spectra |
| EPIC - | 3,519 data sheets |
| AMPIC - | N/A |
| CDC - | N/A |

# APPENDIX B

## List of Chemical Data Centers and Compilations Examined

**Science Communication**
Washington, D. C. 200 07

-B-

C 479 Final Report
20 March 1967

List of Chemical Data Centers and Chemical Data Compilations

A. **The Joint Committee on X-Ray Powder Diffraction Data,**
American Society for Testing and Materials
1916 Race Street, Philadelphia, Pennsylvania

1. X-Ray Powder Data File. Organic, and inorganic. Sets 1-5 in Tabular Format, Revised. 1960.
2. ASTM X-Ray Powder Diffraction Card File. Organic, and inorganic. 3 x 5 card format, updated annually.
3. ASTM X-Ray Powder Diffraction Card File. Organic, and inorganic. Keysort format, updated annually.
4. Matthews Coordinate Index for Inorganic Compounds. Optical coincidence, updated annually.
5. The Fink Inorganic Index to the X-Ray Powder Diffraction File. Published annually.
6. KWIK Index for Inorganics. Published annually.
7. List of Alloy-Phased Designations of the X-Ray Diffraction Data File. Inorganic. Sections 1-12. 1963.
8. X-Ray Powder Diffraction Data File. Organic, and inorganic. Sets 1-16, magnetic tape. Updated annually.
9. The Hanawalt-Davey Index to the X-Ray Powder Diffraction File. Organic, and inorganic. Published annually.
10. The IBM-Tab (Hollerith) Card System, An Index of the ASTM X-Ray Powder Diffraction Data File. Organic, and inorganic. Updated annually.

B. American Society for Testing and Materials (ASTM)
1916 Race Street, Philadelphia, Pennsylvania
**The Wyandotte-ASTM Spectral Data Project**
Wyandotte Chemicals Corporation
Wyandotte, Michigan

1. Mass Spectrometry, Data and Name File. Hollerith cards. 1963.
2. Mass Spectrometry, Data and Chemical Structure File. Hollerith cards. 1963.
3. Molecular Formula List of Compounds, Names and References to Published Infrared Spectra. Hollerith cards. Updated annually.
4. The Wyandotte-ASTM Punch Card Index for Infrared Spectra. Includes data on near infrared, 0.7-3.5 microns, infrared, 2-16 microns, and far infrared, 11-36 microns. Hollerith cards. Updated annually.

5. The Wyandotte-ASTM Punch Card Project for Ultraviolet and Visible Absorption Spectra. 200-420 mμ coding resolution for ultraviolet, 350-845 mμ coding resolution for visible spectra. Hollerith card format. Updated annually.
6. Special Technical Publication 358. Serial Number List of Compound Names and References to Published Infrared Spectra. ASTM 1963.
7. Special Technical Publication 356. The Index of Mass Spectral Data. ASTM 1963.
8. Special Technical Publication 357. Molecular Formula List of Compound Names and References to Published Ultraviolet and Visible Spectra. ASTM 1963.
9. Special Technical Publication 331, and Supplement. The Molecular Formula List of Compound Names and References to Published Infrared Spectra. ASTM 1962.
10. Special Technical Publication 329, and Supplement. Coder for Periodical Titles. ASTM 1964.

C. The University of California, Berkeley, California
**Thermodynamic Properties of Metals and Alloys Data Center**

1. Selected Values of Thermodynamic Properties of Metals and Alloys. John Wiley, 1963.
2. Selected Values of Thermodynamic Properties of Metals and Alloys. Issued in loose-leaf form as supplements.

D. Texas A & M University, College Station, Texas
**Thermodynamics Research Center (formerly MCA) Projects on Properties of Chemical Compounds**

**American Petroleum Institute Project 44**

1. Selected Values of Properties of Hydrocarbons and Related Compounds. Updated semi-annually.
2. The API Catalog of Infrared Spectral Data.
3. The API Catalog of Ultraviolet Spectral Data.
4. The API Catalog of Raman Spectral Data.
5. The API Catalog of Mass Spectral Data.
6. The API Catalog of NMR Data.

7. Selected Values of Properties of Chemical Compounds. Thermodynamics Research Center (formerly MCA). Updated semi-annually.
8. TRC Catalog of Infrared Spectral Data.
9. TRC Catalog of Ultraviolet Spectral Data.
10. TRC Catalog of Raman Spectral Data.
11. TRC Catalog of Mass Spectral Data.
12. TRC Catalog of NMR Spectral Data.

E. National Bureau of Standards, Gaithersburg, Maryland
**Selected Values of Thermodynamic Properties of Inorganic Substances Data Center**

1. Selected Values of Thermodynamic Properties of Inorganic Substances. NBS Circular 500. Updated periodically.

F. Purdue University, Lafayette, Indiana
**Thermophysical Properties Research Center** (TPRC)

1. Thermophysical Properties Research Center Data Book. Tabular format. Updated semi-annually.
2. Retrieval Guide to Thermophysical Properties Research Literature. Tabular format. Published yearly.

G. Dow Chemical Company, Midland, Michigan
**Joint Army-Navy-Air Force (JANAF) Thermochemical Tables**

1. JANAF Thermochemical Tables. In tabular format and magnetic tapes. Updated quarterly.

*H. Atomic Energy Commission
Oak Ridge National Laboratory, Oak Ridge, Tennessee
**Atomic and Molecular Processes Information Center**

1. Bibliography on Reaction Processes. Tabular format. Published semi-yearly.

*I. National Bureau of Standards, Boulder, Colorado
**Cryogenic Data Center**

1. Thermodynamic Properties of Helium at Low Temperatures and High Pressures. May 1959.
2. Thermodynamic Properties of Helium from 3-300°K, Between .5 and 100 Atmospheres. Technical Note #154, 1962.

3. Thermodynamic Properties of Helium from 6-540° R, Between 10 and 1500 PSIA. Technical Note #154a. January 1962.

4. The Solubility of Helium in Liquid Hydrogen. The Journal of Chemistry and Physics, Volume 40. April 1964.

5. A Tabulation of the Thermodynamic Properties of Normal Hydrogen from Low Temperatures to 300°K, and from 1 to 100 Atmospheres. Technical Note #120. November 1961.

6. A Tabulation of the Thermodynamic Properties of Normal Hydrogen from Low Temperatures to 540°R, and from 10 to 1500 PSIA. Technical Note #120a. June 1962.

7. A Compilation and Correlation of the P-V-T Data of Normal Hydrogen from Saturated Liquid to 80°K. Advances in Cryogenic Engineering, Volume 5 (Procedures of the 1959 Cryogenic Engineering Conference. Plenum Press, Inc. 1960).

8. The Densities of Saturated Liquid Hydrogen. Cryogenics #2, volume 2. December 1961.

9. Compilation of the Physical Equilibria and Related Properties of the Hydrogen-Helium System.

10. Thermodynamic Properties of Parahydrogen from 1-22°K. Technical Report #1.

11. Specific Heat of Saturated Liquid Parahydrogen from 15-32°K. Cryogenics #2, volume 5. September 1962.

12. Vapor Pressure of 20°K Equilibrium Hydrogen. Cryogenics #2, volume 4. 1962.

13. Specific Heat at Constant Volume of Parahydrogen at Temperatures from 15-90°K and Pressures to 340 ATM. Cryogenics #2, volume 6. December 1962.

14. Pressure-Density Temperature Relations of Fluid Parahydrogen from 15-100°K at Pressures to 350 Atmospheres. Technical Note #67a. Number 2. April 1963.

15. The Orthobaric Densities of Parahydrogen, Derived Heats of Vaporization and Critical Constants. Cryogenics #2. March 1963.

16. Thermodynamic and Related Properties of Parahydrogen from the Triple Point to 100°K at Pressures to 340 Atmospheres. Monograph #94. August 1965.

17. Refractive Index Dispersion of Liquid Hydrogen. Technical Note #323. September 1965.

18. Low Temperature Transfer Properties of Commercial Metals and Alloys. Journal of Applied Physics, Volume 31, #3. March 1960.
19. Low Temperature Transfer Properties of Copper and Its Dilute Alloys, Pure Copper, Annealed and Cold-drawn. Physics Review, #115, #3. July 1959.
20. Low Temperature Transport Properties of Commercial Metals and Alloys. IV Reactor Grade BE, MO and W. Journal of Applied Physics, volume 31, #7. July 1960.
21. Mechanical Properties of Structural Materials at Low Temperatures: a Compilation from the Literature. Monograph #13. June 1960.
22. Low Temperature Transport Properties of Commercial Metals and Alloys. III Gold-Cobalt. Journal of Applied Physics, volume 31. March 1960.
23. Thermal Expansion of Technical Solids at Low Temperatures: a Compilation from the Literature. Monograph #29. May 1961.
24. Specific Heats and Enthalpies of Technical Solids at Low Temperatures. Monograph #21. October 1960.

*J. Hughes Aircraft Company, Culver City, California
**Electrical-Electronic Properties Information Center** (EPIC)

1. Cadmium Oxide (DS149, M. Neuberger, June 1966).
2. Bismuth Telluride-Bismuth Selenide Systems (DS147, M. Neuberger, January 1966).
3. Zinc Sulfide (DS135, D. B. Carter, November 1966).

State-of-the-Art Reports:

4. Polimide Plastics, A State-of-the-Art Report. (S8, John Milek, October 1965)
5. Niobium Alloys and Compounds (DS148, Donald L. Grisby, January 28, 1966).

* Those Chemical Data Compilations in monograph form were analyzed qualitatively.

**Science Communication**
Washington, D. C. 200 07

C-478 Final Report
20 March 1967

# APPENDIX C

## Description and Flow Diagrams for the Data Compilation Activities

Chemical Thermodynamic Data Group, Thermochemistry Section, Physical Chemistry Division, Institute of Basic Standards, National Bureau of Standards (Circular 500).

The National Bureau of Standards Thermochemistry Section collects, correlates and evaluates data on chemical thermodynamic properties of all inorganic substances and organic substances having molecules containing not more than two carbon atoms. The tables contain values, where known, of the enthalpy and Gibbs energy of formation, enthalpy, entropy and heat capacity at 298.15°K (25°C) and the enthalpy of formation at 0°K. The Standard Order of Arrangement (the principle of latest position) is used to list the compounds, so that in a given element table, all the compounds of that elements will be grouped together.

Their reference source file consists primarily of technical reports published in journals, periodicals and government reports. They make technical evaluations regarding the accuracy, quality and significance of information that has been introduced into the system. They prepare state-of-the-art reviews, correlate information and data, and prepare information on selected classes of compounds for publication. All data and values are critically evaluated and correlated to obtain the best values.

The Center also provides answers to technical questions and inquiries. These references and their principal measured values are extracted on to 3 x 5 cards and entered into a large file which, within our system concept, combines the data source file index and the data file. Upon the decision to compile or update values for a substance-property field, relevant references and data can then be retrieved from the file, which is organized on the basis of substance and property coverage.

Circular 500

Chemical Thermodynamic Data Group
Thermochemistry Section, Physical Chemistry Division
Institute of Basic Standards
National Bureau of Standards

## Physical Arrangement of Data Source Files (Library)

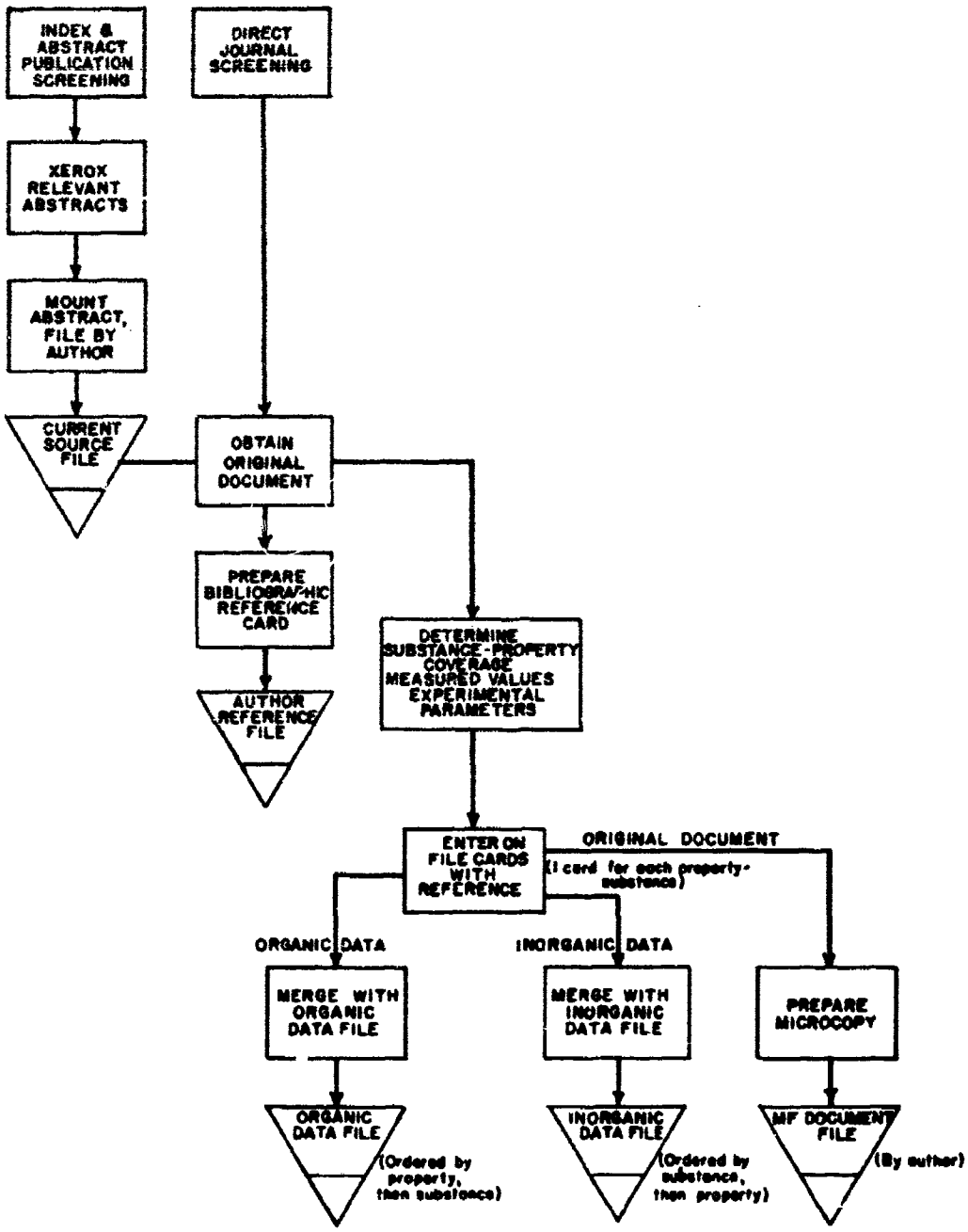
The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | Document File | Author-Reference File | Inorganic Data File | Organic Data File | Selected Data File | Current Source File | Work Files |
|---|---|---|---|---|---|---|---|
| **Recording Media** | 5 x 8 microcopy + tabular | 3 x 5 cards | 3 x 5 cards | 3 x 5 cards | 3 x 5 cards | *3 x 5 cards | 8 1/2 x 11 sheets, calculators, documentation in file folders |
| **File Element** | | | | | | | |
| Source Document | N | | | | | | |
| Document Abstracts | | | | | | N | |
| Personal Author | P | P | N | N | | S | |
| Title | | N | | | | | |
| Journal Source | | N | N | N | | P | |
| Type of Source Document | | | | | | | |
| Date of Publication | | N | N | N | | S | |
| Accession Number | | | | | | | |
| Substance Identified | | | P | S | P | N | P |
| Property Identified | | | S | P | | N | |
| Type of Measurement | | | N | N | | | |
| Experimental Parameters | | | N | N | | | |
| Evaluated Data | | | | | | | |
| State of Substance | | | | | N | | |
| No. of "C" atoms in Substance | | | | S | S | | |
| Chemical Abstract Number | | | | | | N | |

* There is also an identically constructed file of not current sources, unabstracted.

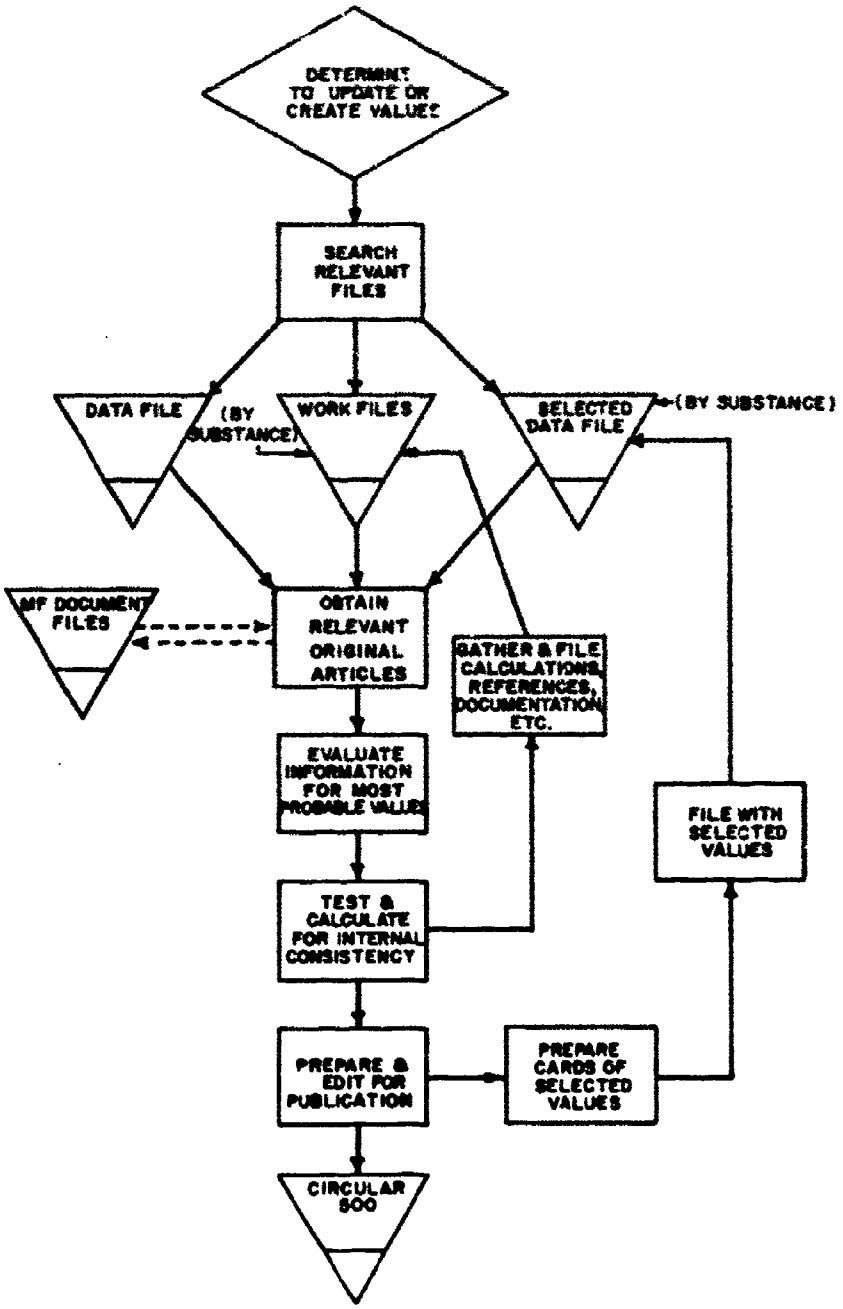
NBS Circular 500 #1

Construction and Maintenance of Data Source and Data File



NBS Circular 500 #2

Searching of Files and Preparation of Selected Values



**Science Communication**
Washington, D. C. 200 07



Joint Army-Navy-Air Force (JANAF) Thermochemical Tables, Dow Chemical Company, Midland, Michigan.

The preparation of the JANAF Thermochemical Tables was initiated as part of a large experimental program carried on at several laboratories. In the early phases, these laboratories, under the coordination of the JANAF Thermochemical Panel, contributed data which were issued without critical evaluation. Now most new data is screened from the open literature, principally Chemical Abstracts, and is all critically evaluated prior to dissemination.

Substance coverage of this data compilation activity is confined to the light elements (first two periods of the periodic table) plus some selected heavy elements. Approximately one-fourth of the elements in the periodic table are covered. Properties coverage includes standard thermodynamic functions as a function of temperature (e.g., heat capacity, entropy, Gibbs free energy and heat of formation).

The principal users of the compilation are specialists who calculate performance of rocket fuels.

The total number of data sheets in the compilation is 950. Approximately 200 data sheets are issued annually, fifty each quarter. Of the 200 issued annually, 100 are data sheets for chemical substances not previously covered, and 100 are revisions of data sheets previously issued. The activity maintains a data source index file, a data source file and a data file. The distribution of effort among the operations performed by the activity is concentrated in the data evaluation operation (approximately 70% of the total). Automated methods are used both to screen the literature for location of source documents and to compute the tables of thermodynamic functions. The use of ADP methods has recently, on a trial basis, been extended to printing of the tables of functions.

## Joint Army-Navy-Air Force (JANAF) Thermochemical Tables

Dow Chemical Company
Midland, Michigan

### Physical Arrangement of Data Source Files (Library)

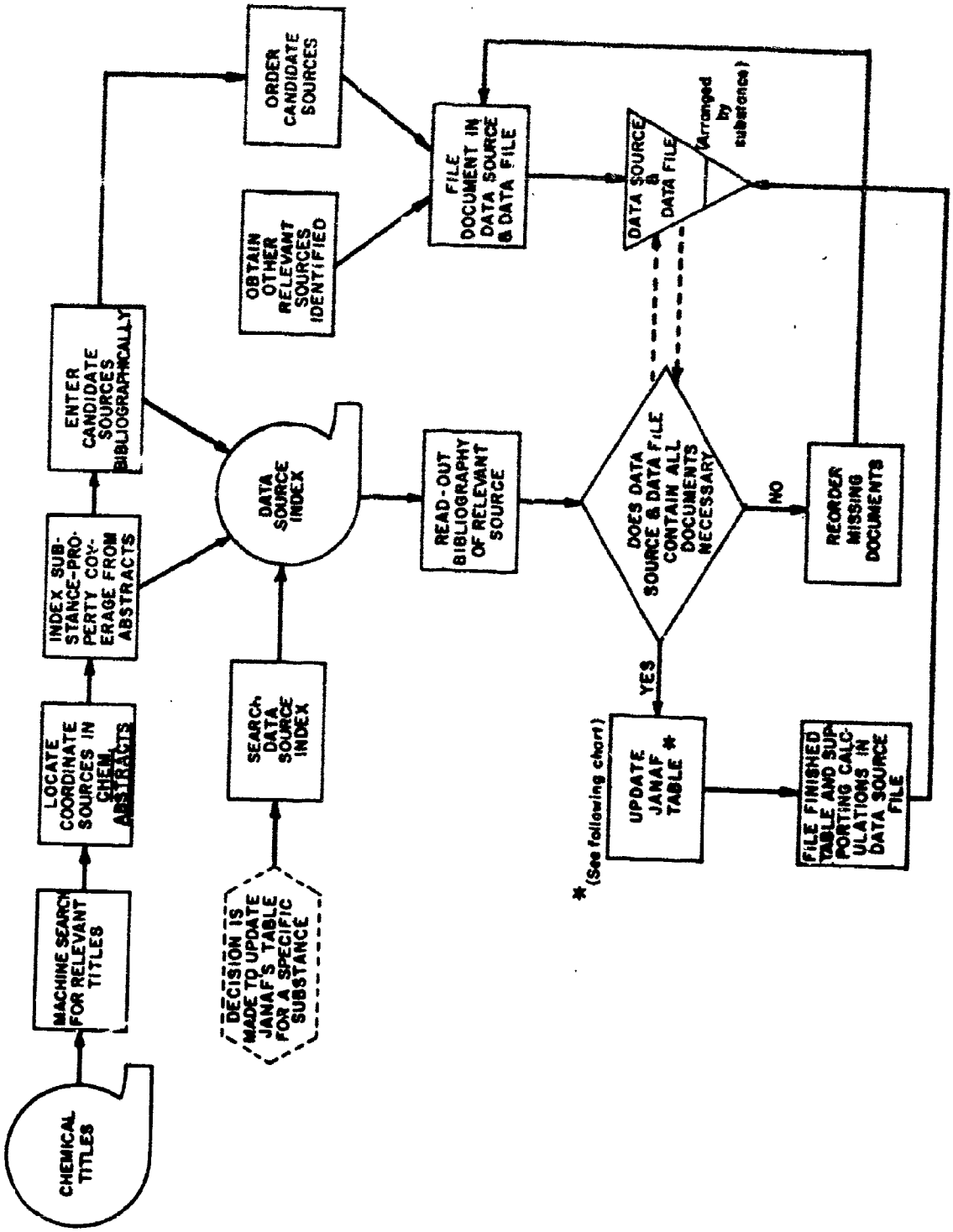
The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | 1925-1962 Source Doc. Index File | 1962-1966 Source Doc. Index File | Data Source Index File | Data Source and Data File | Basic Data File |
|---|---|---|---|---|---|
| Recording Media | 3 x 5 cards | mag. tape | mag. tape | hardcopy | IBM punched cards |
| File Element | | | | | |
| Source Document | | | | N | |
| Document Abstracts | | | | N | |
| Personal Author | N | S | | | |
| Journal Source (coden abbreviation) | N | S | | | |
| Type of Source Document | | | | | |
| Date of Publication | N | S | | | |
| Accession Number | P | P | S | | |
| Substance Identified | N | | P | P | P |
| Property Identified | N | | P | | S |
| Type of Measurement | | | | | |
| Data Extracts | | | | | N |
| Chemical Abstracts Ref's. | | N | | | |
| Form of Document | (Not yet implemented) | | | | |

Combined file (Data Source Index)

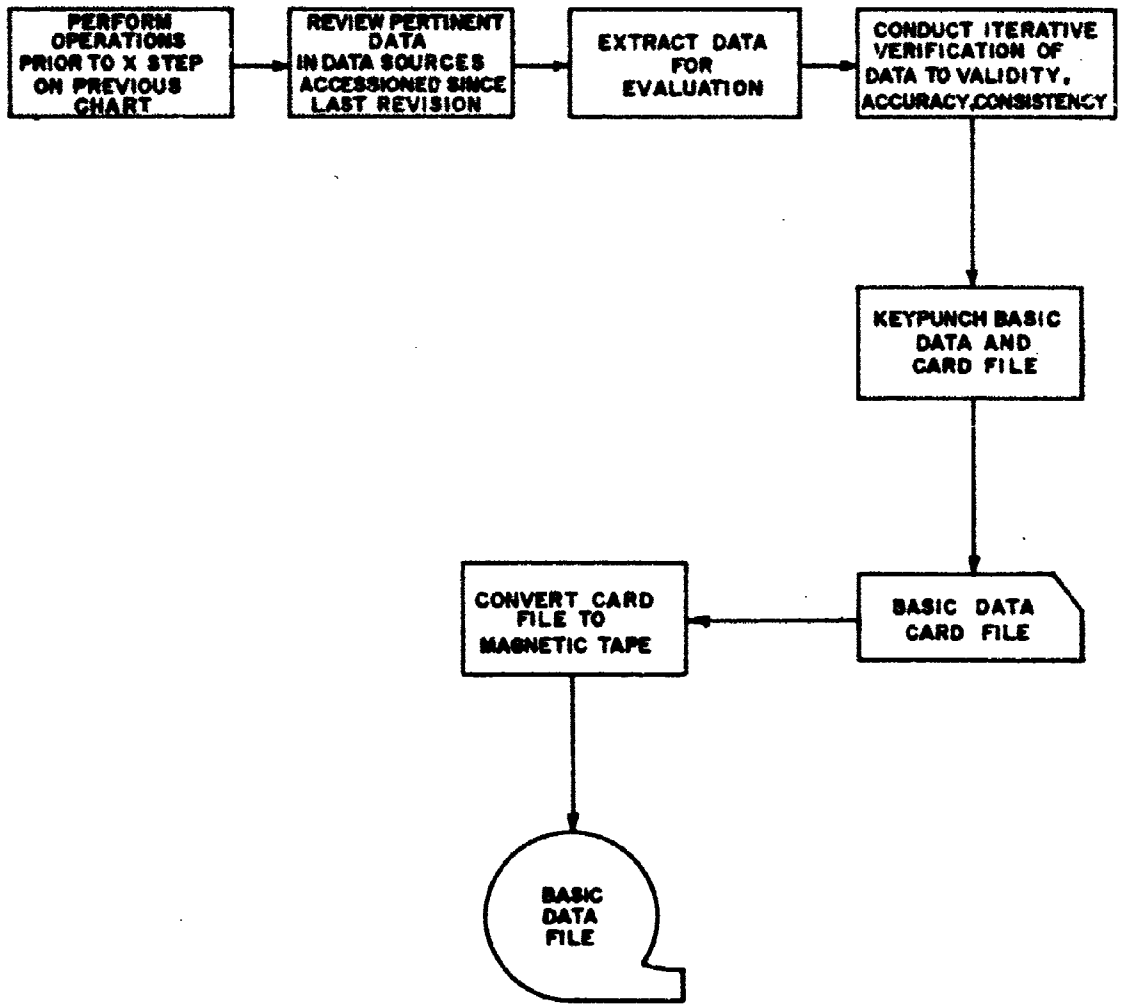
JANAF #1

Construction and Maintenance of Data Source and Data File



## JANAF #2

### Construction and Maintenance of Basic Data File



Thermodynamic Properties of Metals and Alloys Data Center, University of California, Berkeley, California.

The Thermodynamic Properties Data Center compiles data on all substances that conduct electricity electronically, which includes metals and alloys. They evaluate all reported values to produce critical "best estimate" thermodynamic values for all standard thermodynamic functions, thermal functions, vapor pressures, activity coefficients and partial molar quantities.

To support this critical evaluation, the literature is screened through direct journal scanning and abstract journal screening. Relevant sources (coverage from 1910 to date) are indexed and entered in the project's optical coincidence retrieval index, to be retrieved and evaluated upon the decision to compile or update the values for an element or alloy. Selected values were published in book form in 1963 as Selected Values of Thermodynamic Properties of Metals and Alloys and are now updated and expanded by frequent loose-leaf supplements.

## Thermodynamic Properties of Metals and Alloys Data Center

University of California
Berkeley, California

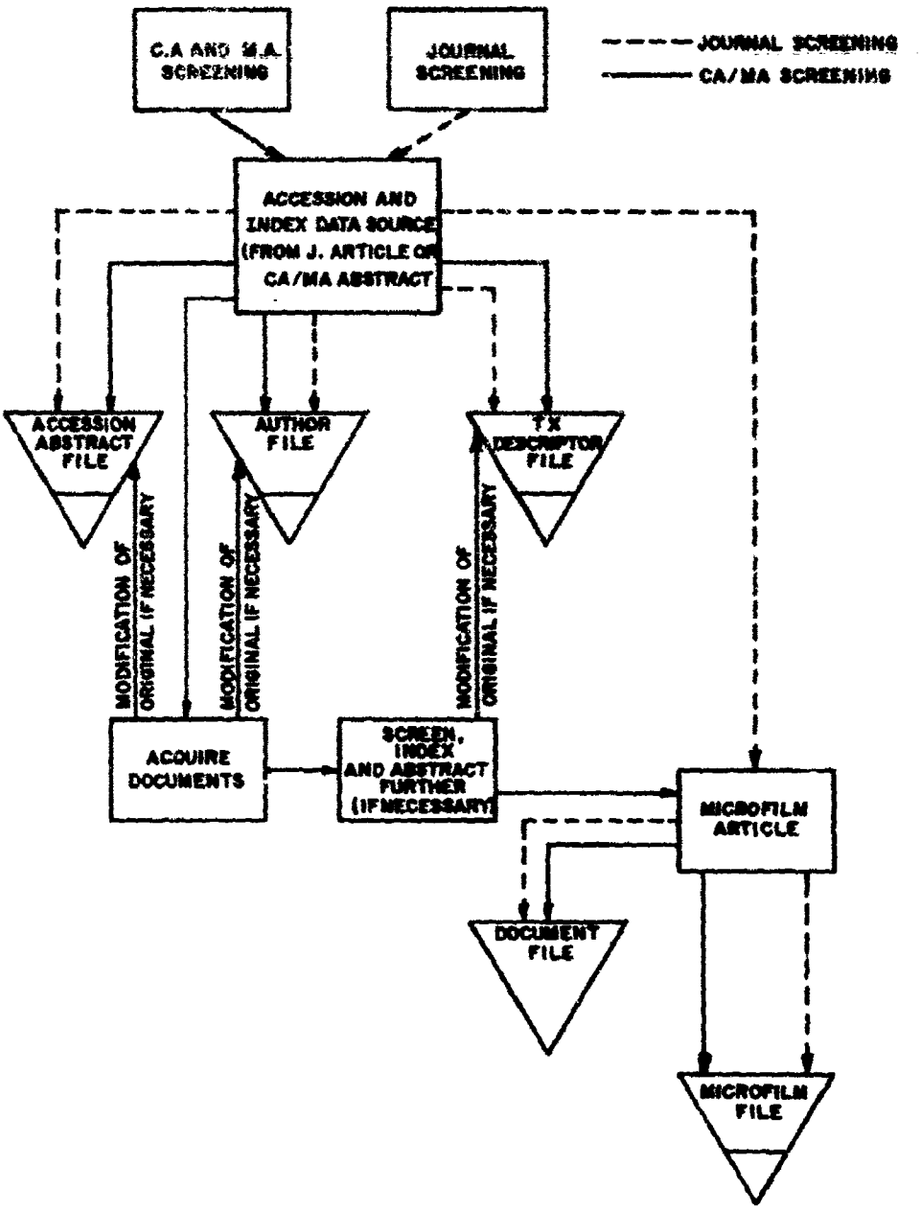
### Physical Arrangement of Data Source Files (Library)

The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | Document File | Document File | Accession Abstract File | Author File | Descriptor File |
|---|---|---|---|---|---|
| Recording Media | Hardcopy | Microcopy | Cards (5 x 7) | Cards | Termatrex |
| File Element | | | | | |
| Source Document | N | N | | | |
| Document Abstracts | | | N | | |
| Personal Author | P | | N | P | P |
| Journal Source | | | N | N | |
| Type of Source Document | | | N | N | |
| Date of Publication | | | N | N | P |
| Accession Number | | P | P | N | N |
| Substance Identified | | | | | P |
| Property Identified | | | | | P |
| Type of Measurement | | | | | P |
| Type of Alloy | | | | | P |
| Title | | | N | N | |

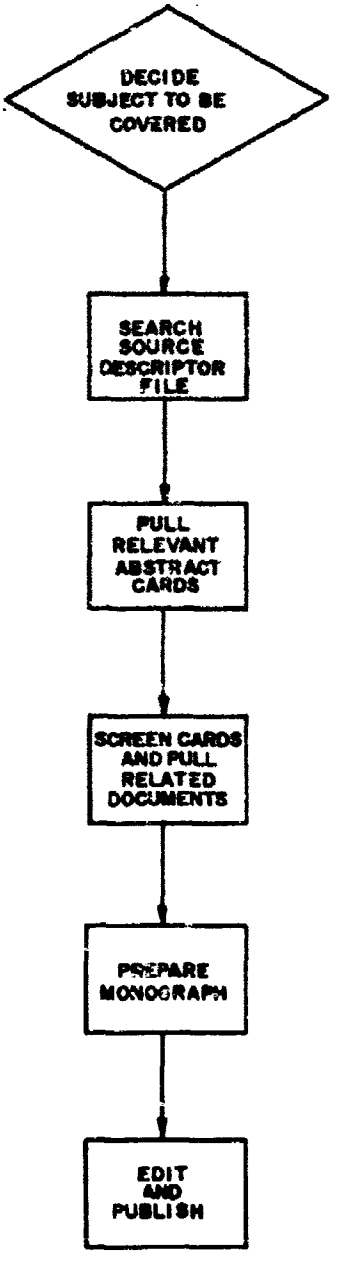
Thermodynamic Properties of Metals and Alloys Data Center #1

Construction and Maintenance of Data Source File



Thermodynamic Properties of Metals and Alloys Data Center #2

Searching of Files and Preparing Output



Cryogenic Data Center, Cryogenic Engineering Laboratory, National Bureau of Standards, Boulder, Colorado.

The Cryogenic Data Center is engaged in the critical evaluation and compilation of data on thermodynamic, transport and other thermophysical properties for the principal cryogenic fluids and common mixtures of these fluids used at low temperatures. The scope of this compilation activity also includes the properties of metallic elements, selected alloys, and element dielectrics. The thermodynamic properties of fluids include pressure-volume-temperature, vapor pressure, latent heat, saturation densities, isothermal compressibility, volume expansivity, entropy, enthalpy, internal energy, specific heat and velocity of sound. The transport properties of fluids include thermal conductivity, Prandtl number, thermal diffusion, viscosity, diffusion and coefficients, while other thermophysical properties include dielectric constant, electrical resistivity, magnetic properties, refractive index, surface tension, optical properties and dielectric breakdown.

This group monitors the literature continually for the above program. As specific tasks are undertaken, comprehensive bibliographies are prepared from their magnetic tape subject index of relevant source documents. These bibliographies are sometimes published. Task notebooks are made for the preliminary selection and evaluation of data, and, where feasible, preliminary memoranda sheets listing the better data available are issued. Their final evaluations yield critical review monographs, presenting tables of "best estimate" values.

All cataloging and coding within the data source file is converted to machine-readable form for automated processing. The programs used are for searching, dictionary term identification, catalog tape output, tape updating, etc. They issue a weekly Current Awareness service, custom bibliographies, preliminary data and advice on the thermodynamic and transport properties of cryogenic fluids. Further, they publish cryogenic laboratory publications and reports in tabular form.

Cryogenic Data Center

Cryogenic Engineering Laboratory
National Bureau of Standards
Boulder, Colorado

## Physical Arrangement of Data Source Files (Library)

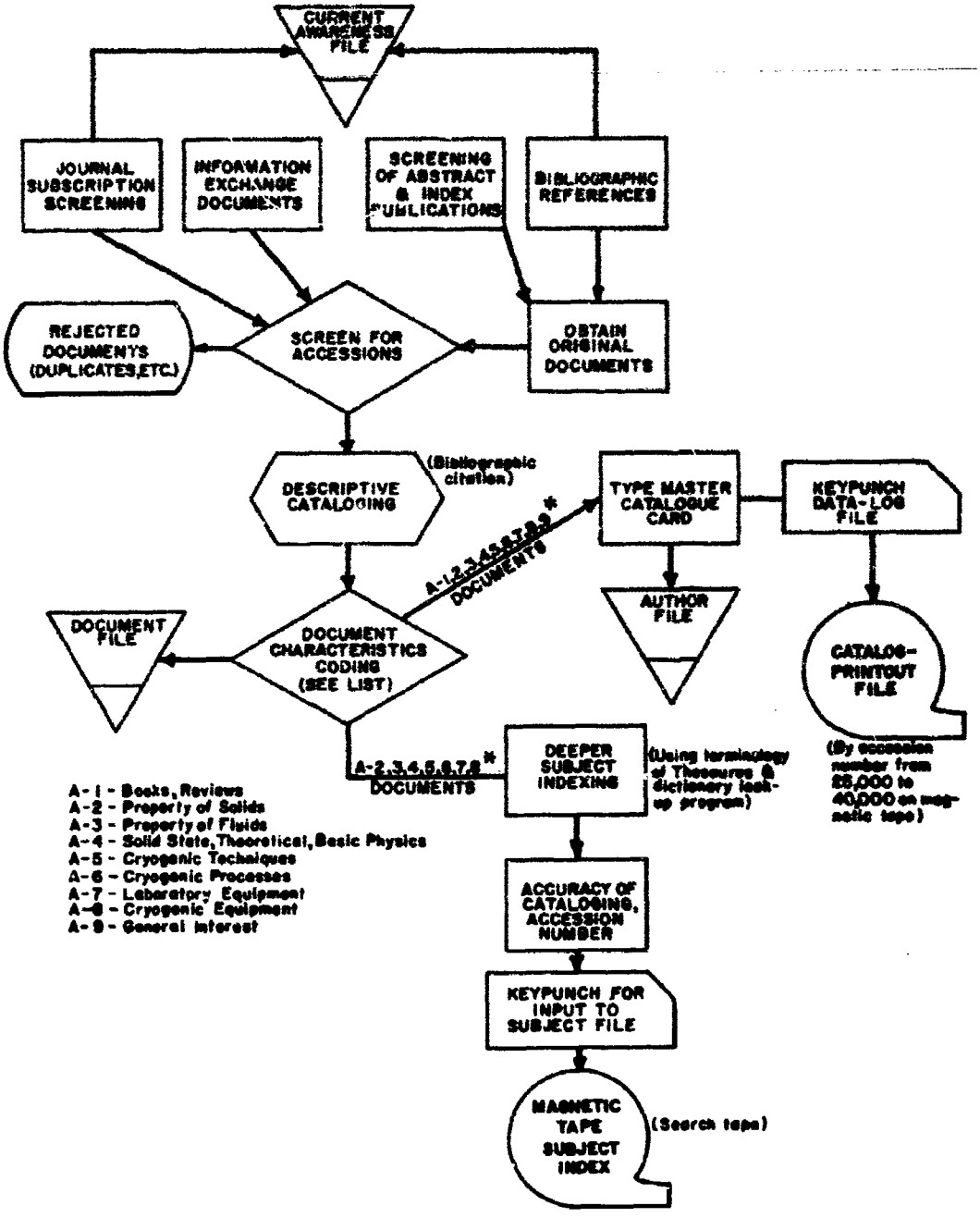
The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file, and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | Document File | Descriptor (Search) File | Author File | Catalog Print-out File |
|---|---|---|---|---|
| Recording Media | Hard copy + microfilm | Magnetic Tape | File cards (master catalog card) | Paper tape - up to 25,000; Magnetic tape |
| File Element | | | | |
| Source Document | N | | | |
| Document Abstracts | | | N | |
| Personal Author | S | | P | N→ author |
| Journal Source | | | N | N~ reference list* |
| Type of Source Document | | | N | N~ coden journal listing* |
| Date of Publication | | | N | N |
| Accession Number | P | N | N | P→ accessions list |
| Substance Identified | | P | N | N |
| Property Identified | | P | N | N |
| Type of Measurement | | | N | N |
| Subject Category | | S | N | N |
| Temperature Range | | | N | |

* Printouts for these lists would be only for the last 15,000 documents, since only they are on magnetic tape (tab cards).

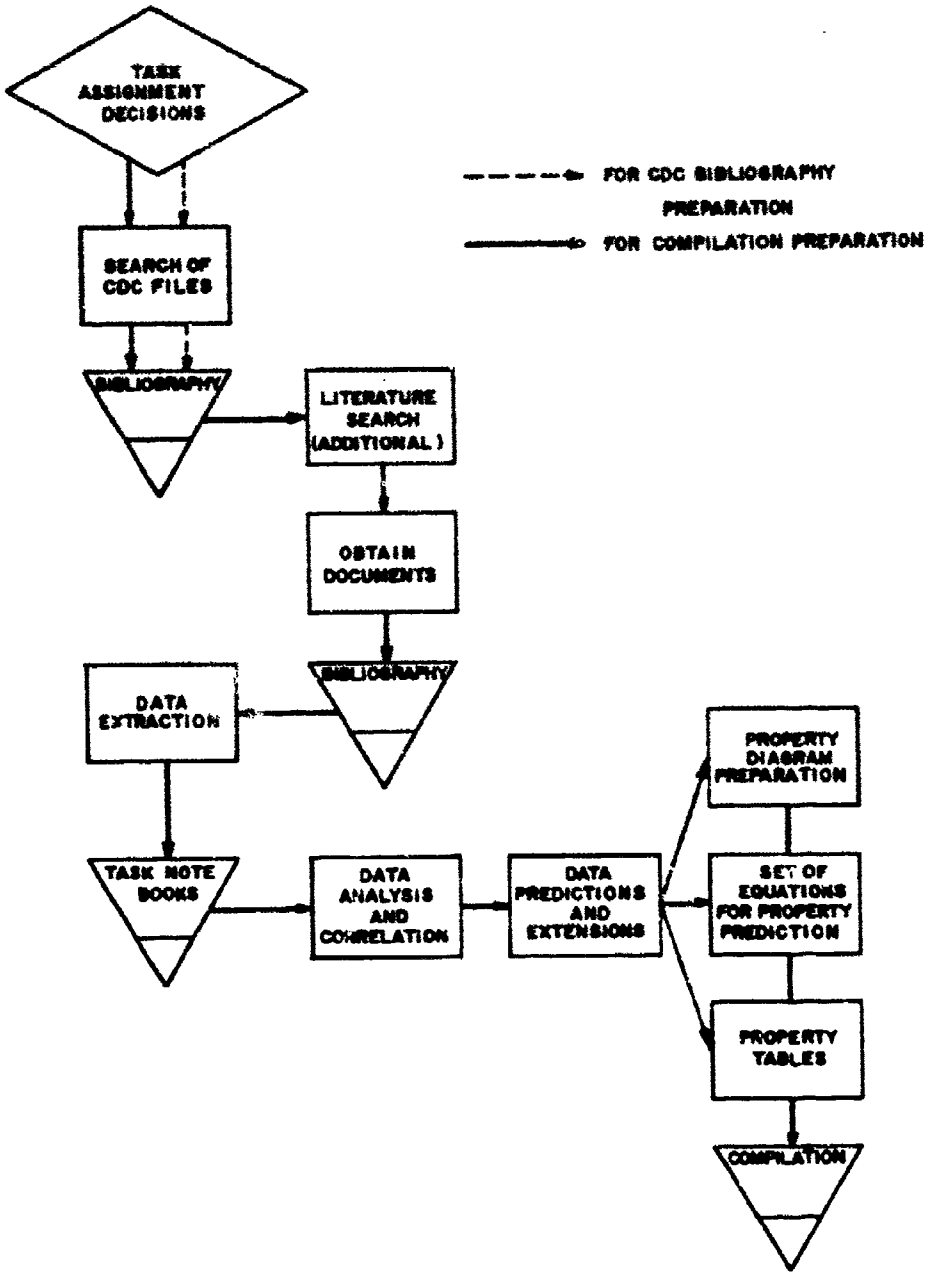
Cryogenic Data Center #1

Construction and Maintenance of Data Source File



Cryogenic Data Center #2

Searching of Files and Preparing Output



Thermophysical Properties Research Center, Purdue University Research Park, Lafayette, Indiana (TPRC).

The principal output of TPRC is the TPRC Data Book on the Thermophysical Properties of Materials. The Data Book presents experimental values from the open literature on thermal conductivity, viscosity, radiative properties, thermal diffusivity and specific heat of materials as comparative curves on a master plot. TPRC develops a curve of recommended values for the property and the material covered over a wide temperature range, from the plot, the sample specifications given and the data points rendered. This work, supported principally by the Air Force, with NSRDS support on selected chemically pure substances, is made available both in the Data Book and in NSRDS reference monographs. All relevant information developed in other data projects undertaken by TPRC for specific purposes, such as their current revision of the Air Force's Handbook on Thermophysical Properties of High Temperature Solid Materials, is incorporated into the Data Book.

The Documentation Division of TPRC supports this effort with nearly 34,000 indexed references, completely covering the literature on thermal conductivity, specific heat, viscosity, emissivity, mass diffusivity, thermal diffusivity and Prandtl number from 1920 to 1965. This nearly-100% coverage will be made available to the public in the forthcoming Volume III of the "Retrieval Guide to Thermophysical Properties Research Literature", which, in essence, will give to its holder access to TPRC's entire literature collection at the time of printing.

The Data Table and Documentation Divisions constitute support for TPRC's long-range goal of documenting and analyzing investigated areas on the thermophysical properties of materials, and then proceeding to experimental and theoretical determination of values for the gaps in knowledge which have been uncovered.

Thermophysical Properties Research Center

Purdue University Research Park
Lafayette, Indiana

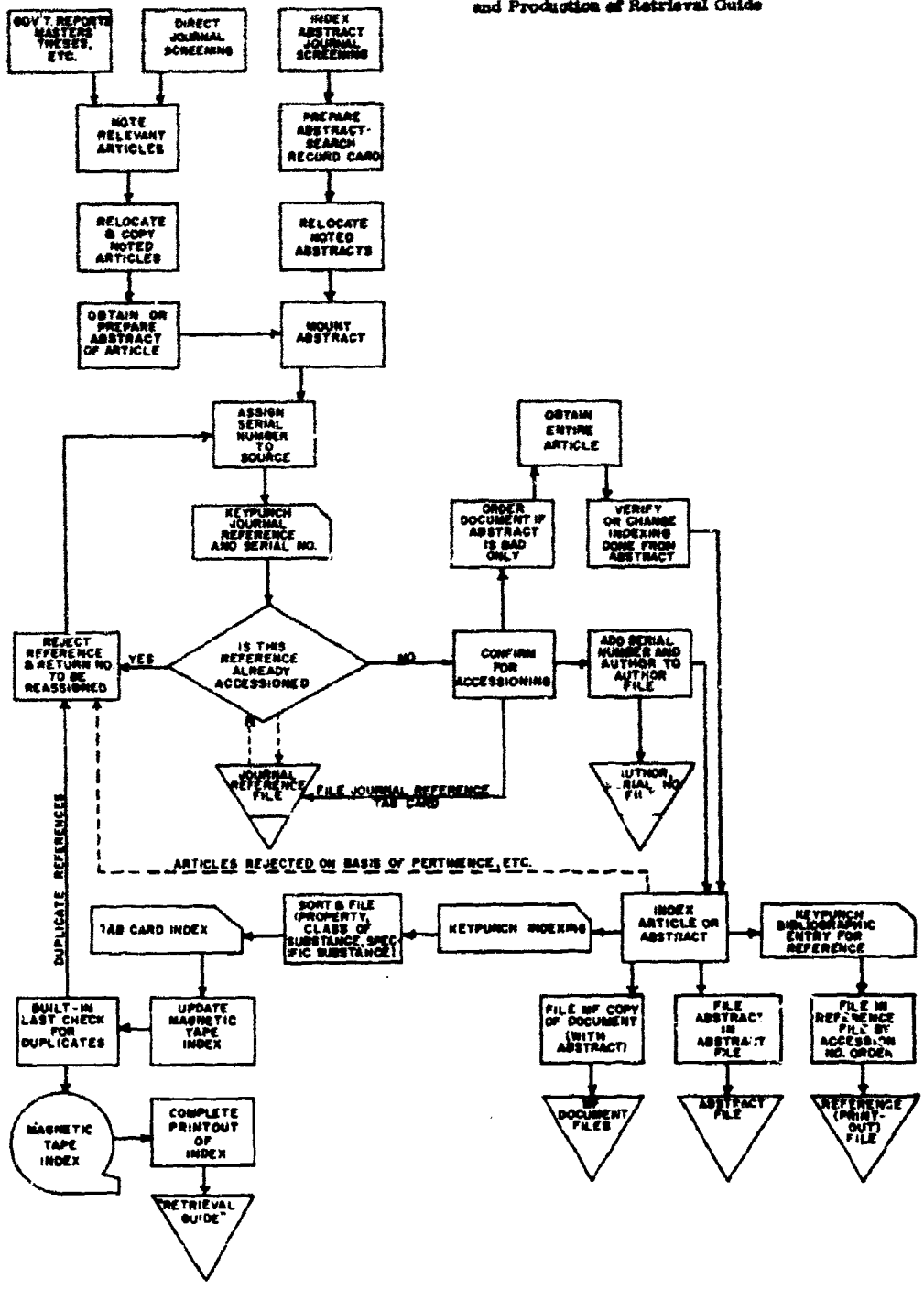
## Physical Arrangement of Data Source Files (Library)

The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | Document File | Abstract File | Descriptor File (retrieval guide) | Reference Printout File | Author File | Journal Reference File |
|---|---|---|---|---|---|---|
| Recording Media | hardcopy being converted to mf. | hardcopy | tab cards/ mag tape | tab cards | tab cards + manual | tab cards |
| File Element | | | | | | |
| Source Document | N | | | | | |
| Document Abstracts | N | N | | | | |
| Personal Author | | N | | N | P | |
| Journal Source | | N | S | N | | P |
| Type of Source Document | | | | | | |
| Date of Publication | | N | S | N | | S |
| Accession Number | P | P | N | P | N | N |
| Substance Identified | | | P | | | |
| Property Identified | | | S (P in future) | | | |
| Type of Measurement | | | | | | |
| Title | | N | | N | | |
| Abstract Identification | | N | | N | | |
| Treatment of Subject | | | S | | | |
| Substance State | | | S | | | |
| Language | | | S | | | |
| Unique Author Number | | | | | N | |

TPRC-Purdue #1

Construction and Maintenance of Data Source File
and Production of Retrieval Guide



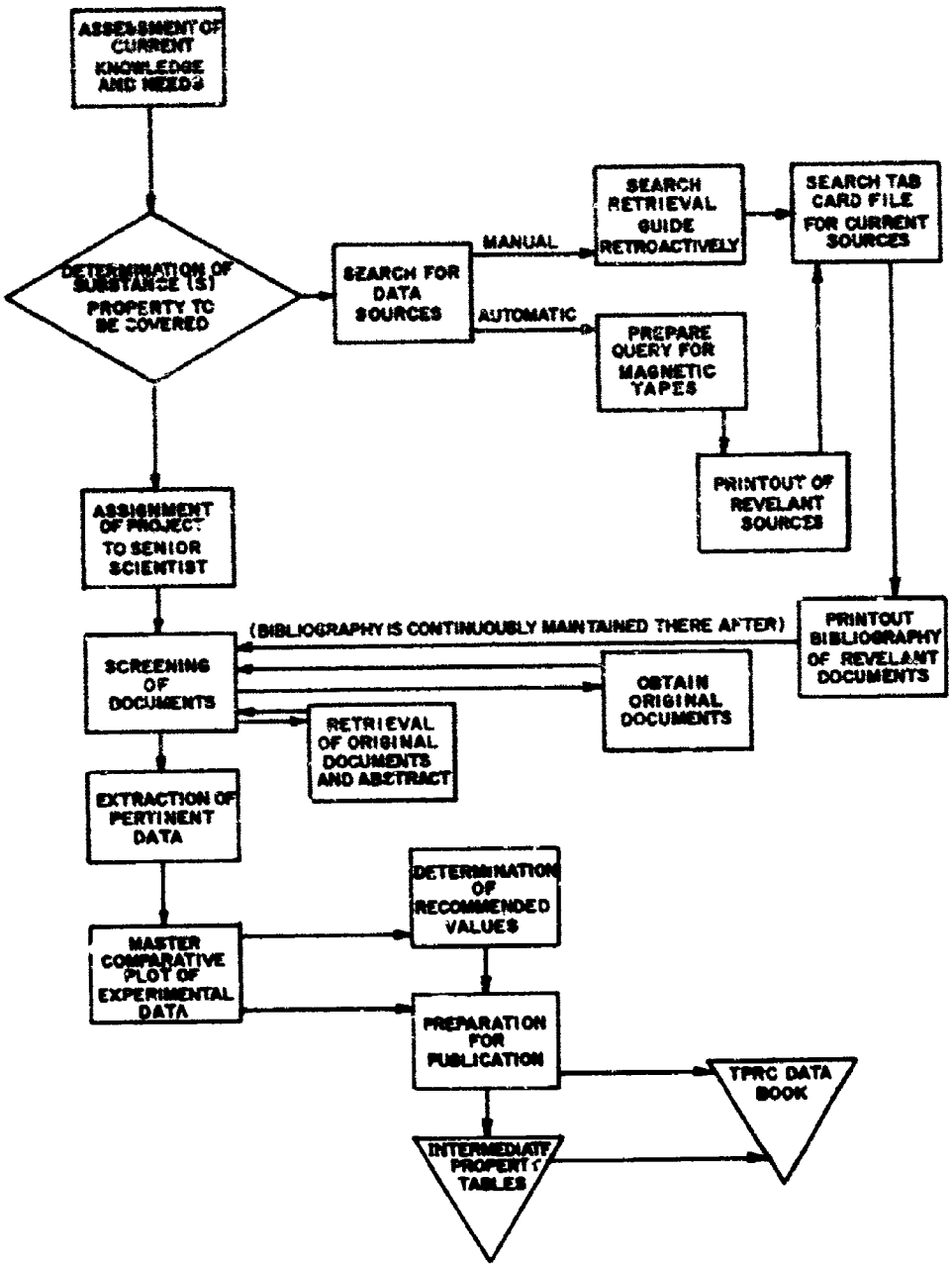
**Science Communication**
Washington, D. C. 200 07



C-478 Final Report
20 March 1967

TPRC-Purdue #2

Searching of Files and Production of Data Book



Thermodynamics Research Center, Texas A & M University, College Station, Texas (TRC).

TRC is the housing organization for the generation and publication of a multiplicity of chemical data compilations. For the American Petroleum Institute, TRC directs API Project 44, to gather, produce internally, evaluate and disseminate from the open literature the "best estimate" values for thermodynamic and thermochemical properties of selected hydrocarbons and for related compounds in cumulative, internally consistent tables. They also prepare and distribute catalogs of Ultraviolet, Infrared, Nuclear Magnetic Resonance, Raman and Mass Spectral data of hydrocarbons and related compounds.

TRC, with support from NSRDS and Texas A & M, compiles the "Selected Values of Properties of Chemical Compounds" (formerly supported by the Manufacturing Chemists' Association) on substances other than those under the scope of API but of interest to the chemical industry. These cumulative tables are presented in the same format as those of API Project 44. Likewise, for substances outside the scope of API Project 44, they distribute catalogs of Ultraviolet, Infrared, Nuclear Magnetic Resonance, Raman and Mass Spectral data.

Original spectra for all the spectral data catalogs are furnished by contributing laboratories. Supporting the compilation of selected values is a massive 3 x 5 card file which, within our system concept, combines the data source file index and data file, and which is organized on the basis of substance and property coverage. References and measured values gleaned from a thorough search of the open literature are entered on cards and merged into the file. The relevant references and data can then be retrieved upon the decision to compile values for a substance-property field. In-house experimental and theoretical research is also conducted for the determination of needed values and the evaluation of experimental values, as well as the calculation of "best estimate" values for publication in the tables of selected values.

In addition to these data compilation efforts, TRC annually contributes a bibliography of the articles on organic substances which have appeared in the open literature of the previous year to the International Union of Pure and Applied Chemistry. For access to this bibliography, a detailed substance-property index is prepared.

Thermodynamics Research Center

Texas A & M University
College Station, Texas

## Physical Arrangement of Data Source Files (Library)

The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file, and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

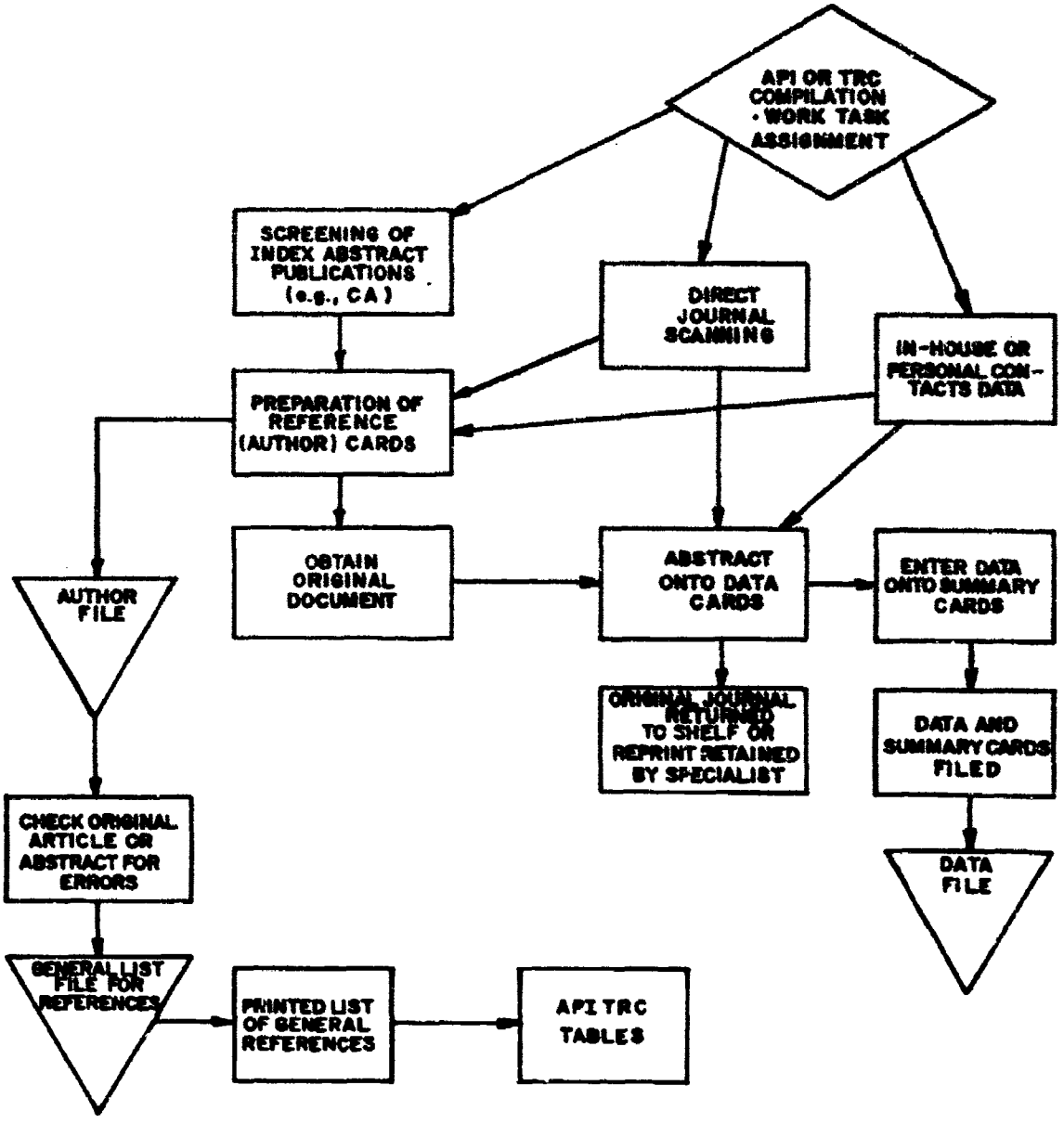
| | Data File | Author File |
|---|---|---|
| Recording Media | 3 x 5 cards | 3 x 5 cards (for IUPAC) |
| File Element | | |
| Source Document | | |
| Document Abstracts (extracts) | N | N |
| Personal Author | N | P (1st author) |
| Journal Source | N | N |
| Type of Source Document | N | N |
| Date of Publication | N* | N |
| Accession Number | | |
| Substance Identified | P** | N |
| Property Identified | S | N |
| Type of Measurement (substance purity, et al.) | N | N |
| Compound Identified | S | N |

* except cards filed chronologically under "Conpound Identified"

** by major classes, e.g., alcohols, et al.

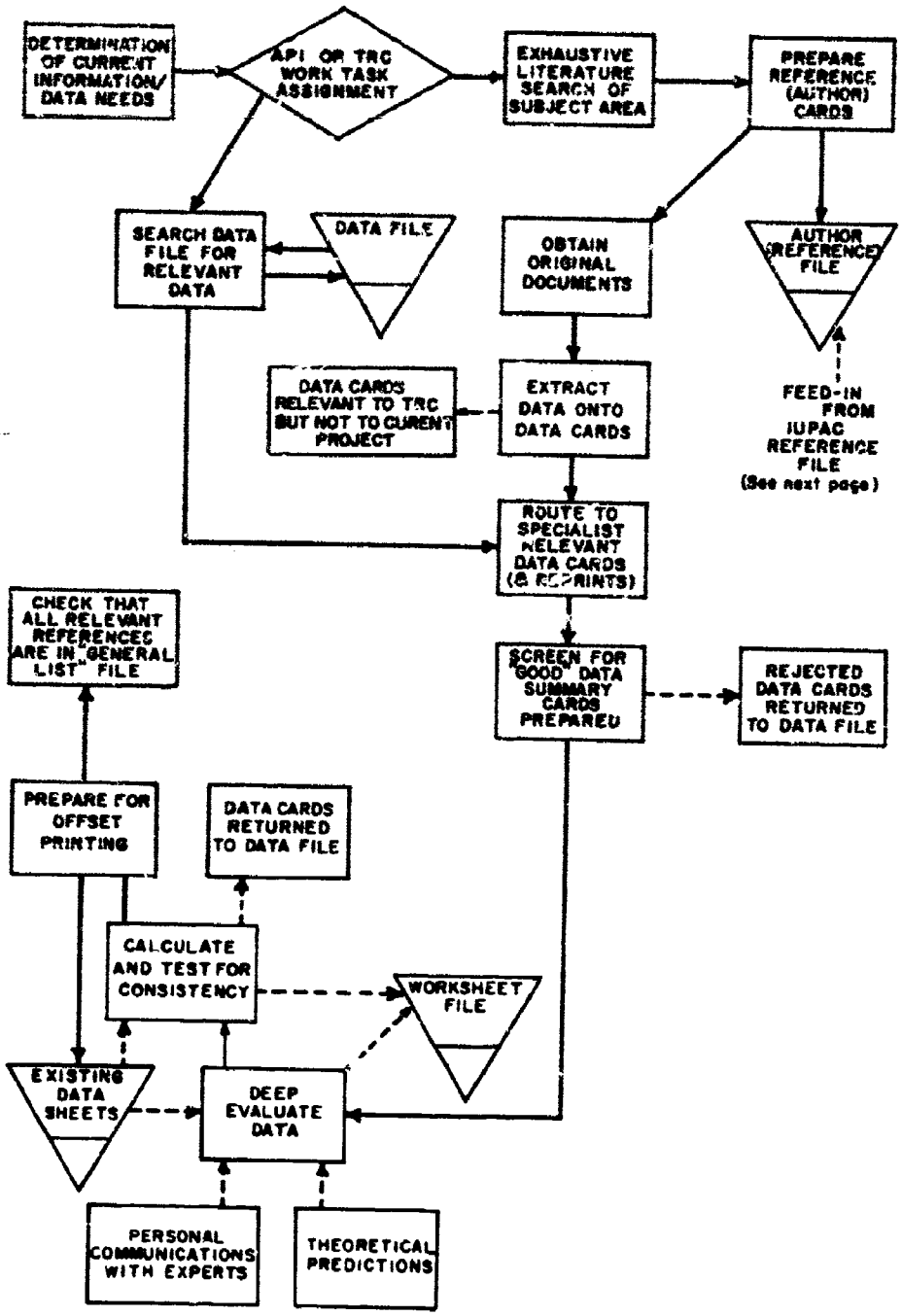
## API - #1

### Construction and Maintenance of Data Source-Data File



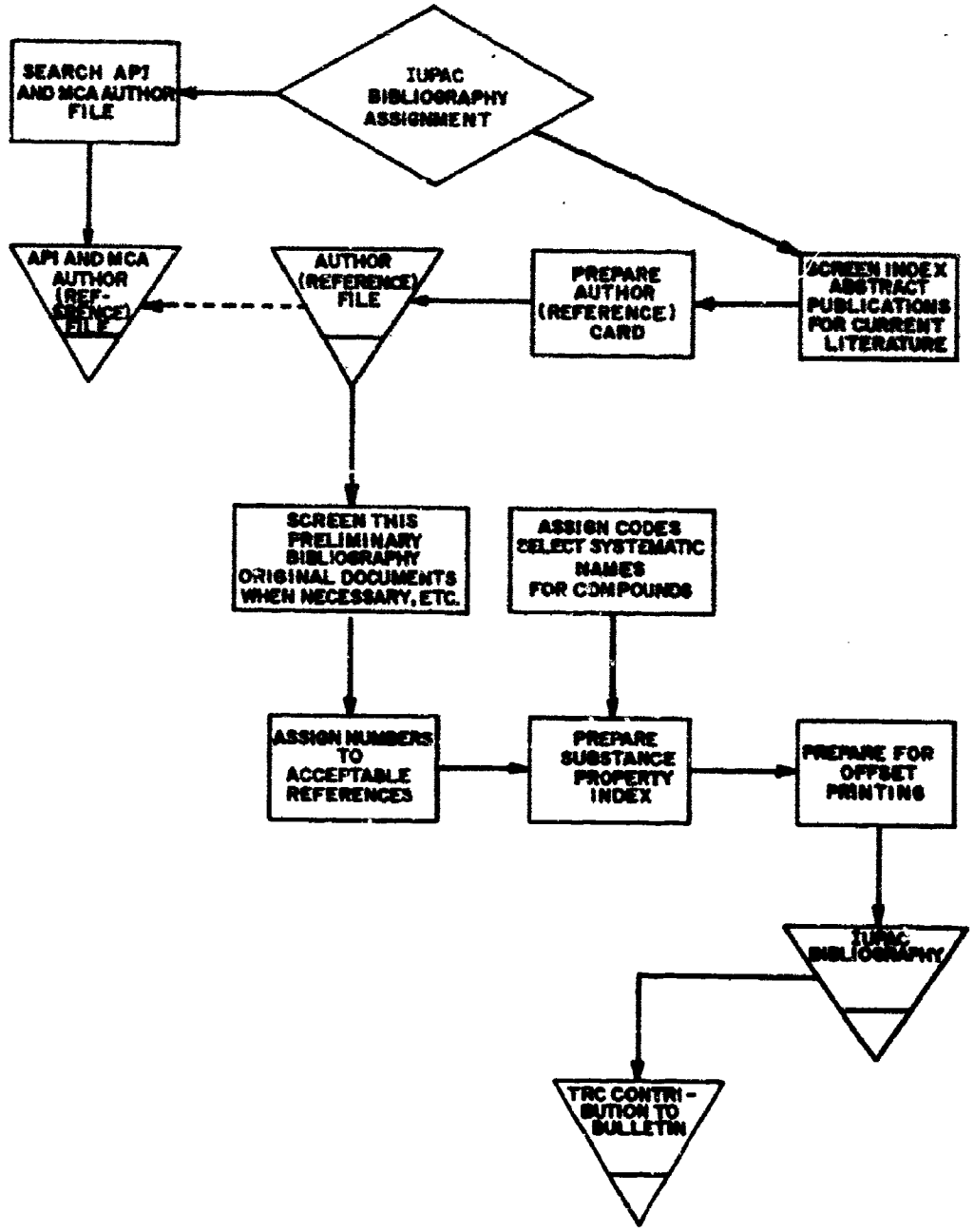
API - #2

Searching of Files and Preparation of Selected Values



API - #3

**Preparation of IUPAC Bibliography**



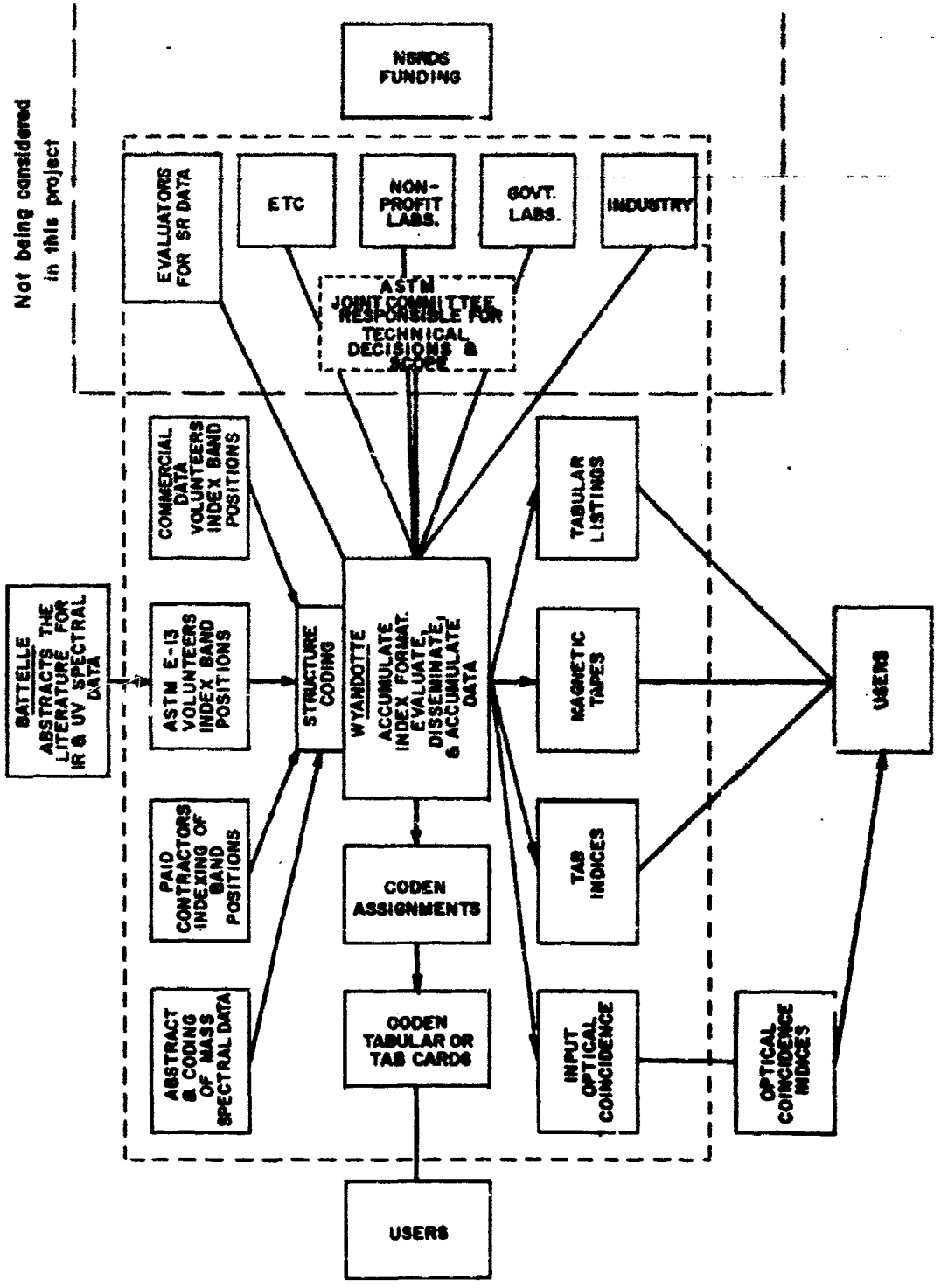
The Wyandotte-ASTM Spectral Data Project, Wyandotte Chemicals Corporation, Wyandotte, Michigan (sponsored by the American Society for Testing and Materials (ASTM), Philadelphia, Pennsylvania).

ASTM prepares and distributes the Wyandotte-ASTM punched card index for searching and correlating Infrared, Ultraviolet and Visible Spectral Absorption data and chemical structure. The system, developed at Wyandotte Chemicals Corporation by Dr. L. E. Kuentzel and subsequently adopted by ASTM Technical Committee E-13 on Absorption Spectroscopy, makes use of Hollerith cards and sorters, and magnetic tape and computers with many different types of output, including optical coincidence, tab card, computer tapes and tabular publications.

The system was designed to facilitate the searching of spectral absorption data for the purpose of matching spectrograms in qualitative analysis and for correlating chemical structure and absorption band positions. A standardized coding system was developed for automated manipulation of the data. The data file includes the spectral position of all important bands of a published spectrogram, the structural details of the chemical compound involved, a semi-molecular formula, and melting and/or boiling point and a reference to the location of the spectrogram in the literature. They attempt to include all published standard spectra including those sold by commercial laboratories.

Wyandotte-ASTM Spectral Data Project

Operational Procedure Flow Diagram



Joint Committee on Chemical Analysis by Powder Diffraction Methods, American Society for Testing and Materials, Philadelphia, Pennsylvania.

The Joint Committee is sponsored by four international cooperating societies: The Institute of Physics (British), The American Crystallographic Society, The National Association of Corrosion Engineers, and The American Society for Testing and Materials.

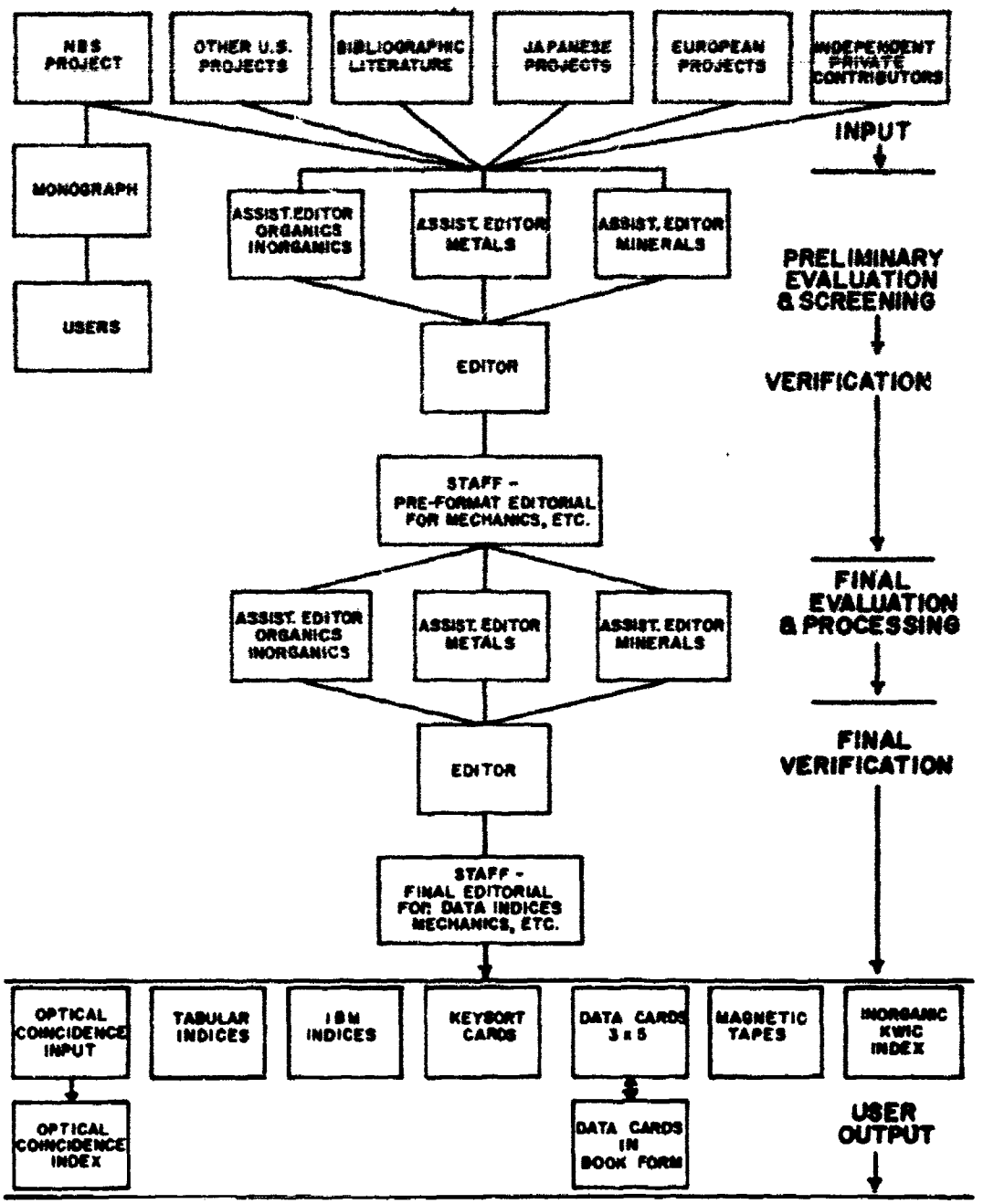
The Joint Committee has two functions: first, to collect diffraction data from all available sources, critically edit this material, publish the selected data in a convenient form, and provide suitable indexes; secondly, to advance the techniques by which this diffraction data can be used for chemical identification. Data are continuously collected for the file in three ways: (1) abstracted from the literature; (2) sponsored projects of the Joint Committee; and (3) patterns sent to the Joint Committee by individual scientists throughout the world.

The Joint Committee sponsors the production of patterns by several groups in the United States, Great Britain, The Netherlands, Israel and Japan, which has resulted in the rapid increase of high-quality patterns, the main factor in growth of the file.

Diffraction analysis identifies a substance by means of atomic arrangement. It is used whenever it is necessary to identify the state or combination of the chemical elements or phases present. The X-ray diffraction cards show numerical values for the powder pattern lines, with intensities, value constants, space group and other crystallographic data, when available. These patterns are presented on three different types of output: plain cards (3 x 5), Keysort cards (4 x 6), and magnetic tapes. They also have indexes that include tabular, books, IBM cards, KWIC indexes, magnetic tapes and optical coincidence.

ASTM-X-Ray Powder Diffraction Project

Operational Procedure Flow Diagram



Electrical-Electronic Properties Information Center, Hughes Aircraft Company, Culver City, California (EPIC).

EPIC searches the open literature for all articles dealing with the electronic properties of materials. When located, such literature is indexed by substance and by one or more of 56 electronic properties. From this data source file and index, a bibliography can be produced that is extensively subject-indexed. Their data compilation efforts are supported by this data source file. EPIC's principal data output is in the form of data sheets. After a material or group of materials is selected for compilation coverage, an assigned specialist reviews all relevant documents, evaluates them for reliability, and then extracts, for re-publication in a concise format, the relevant data content of all reliable articles on the substance and property chosen. At the discretion of the assigned specialist, findings are presented graphically or in a tabular format (one reference per page or a comparison of references on one page). Bound data sheets (in our terminology, monographs) are then issued, giving the electronic properties of a substance or a group of substances. State-of-the-art reports are also prepared by EPIC, presenting experimental data on all properties of selected substances. These reports require searches beyond the immediate EPIC data source files and indexing.

EPIC also produces and distributes large computer-generated bibliographies. These may be as by-products of a data compilation task, or for service as Interim Reports while final data compilation is under way. Such Interim Reports may have some preliminary curves drawn. Specific requests for data and/or references are also answered. Large bibliographies generated in response to such requests are sometimes generally distributed.

Electrical-Electronic Properties Information Center

Hughes Aircraft Company
Culver City, California

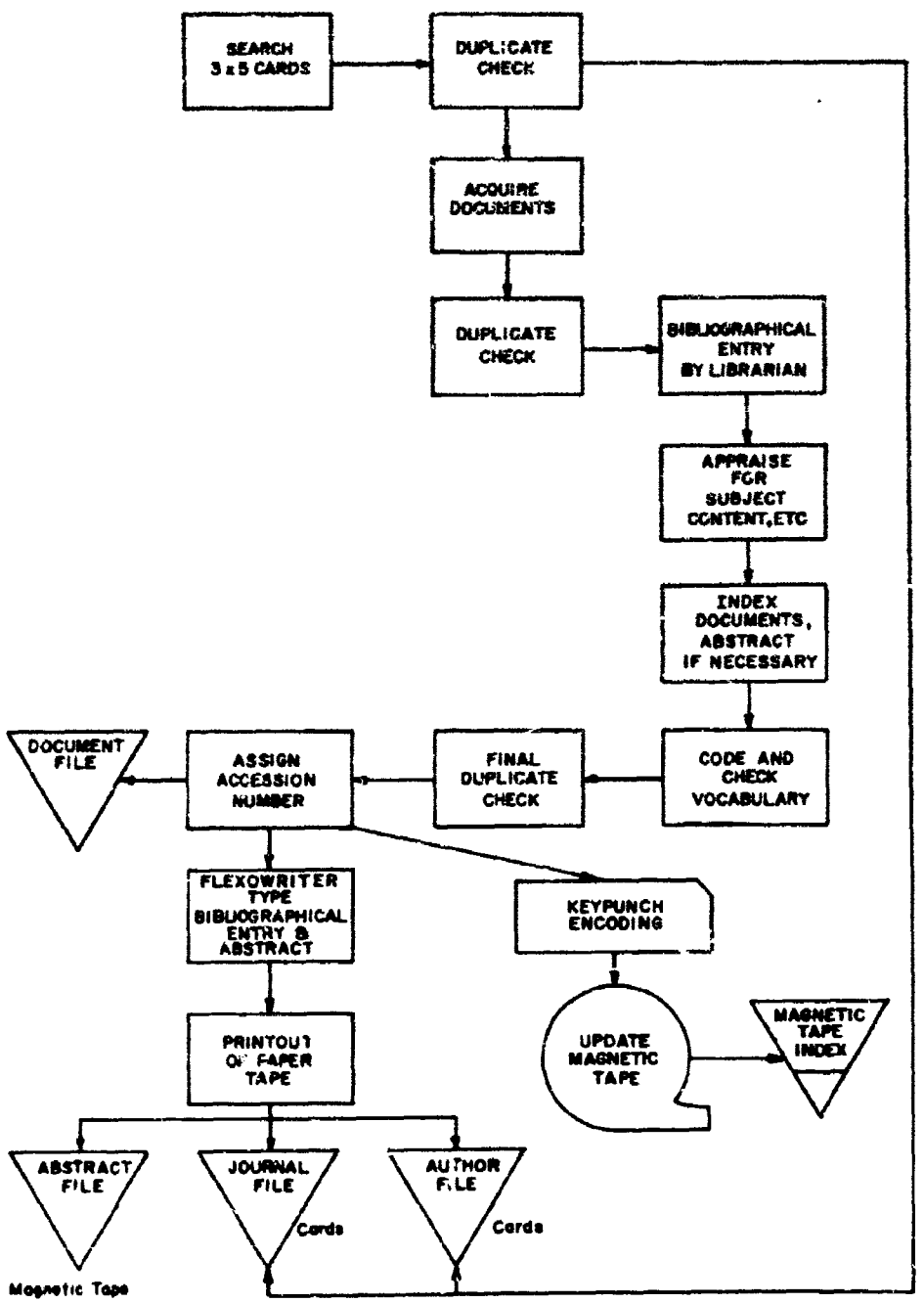
## Physical Arrangement of Data Source Files (Library)

The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | Abstract File | Author File | Journal File | Document File | Descriptor File |
|---|---|---|---|---|---|
| Recording Media | mag. tape | file cards | file cards | hardcopy | mag. tape |
| File Element | | | | | |
| Source Document | | | | N | (can produce three indices or listings for this file) |
| Document Abstracts | N | | | | |
| Personal Author | N | P | N | | |
| Journal Source | N | N | P | | |
| Type of Source Document | N | N | N | | |
| Date of Publication | N | N | N | | |
| Accession Number | P | N | N | P | P |
| Substance Identified | | | | | P |
| Property Identified | | | | | P |
| Type of Measurement | | | | | |
| Title | N | N | N | | |
| Corporate Author | N | S | N | | |

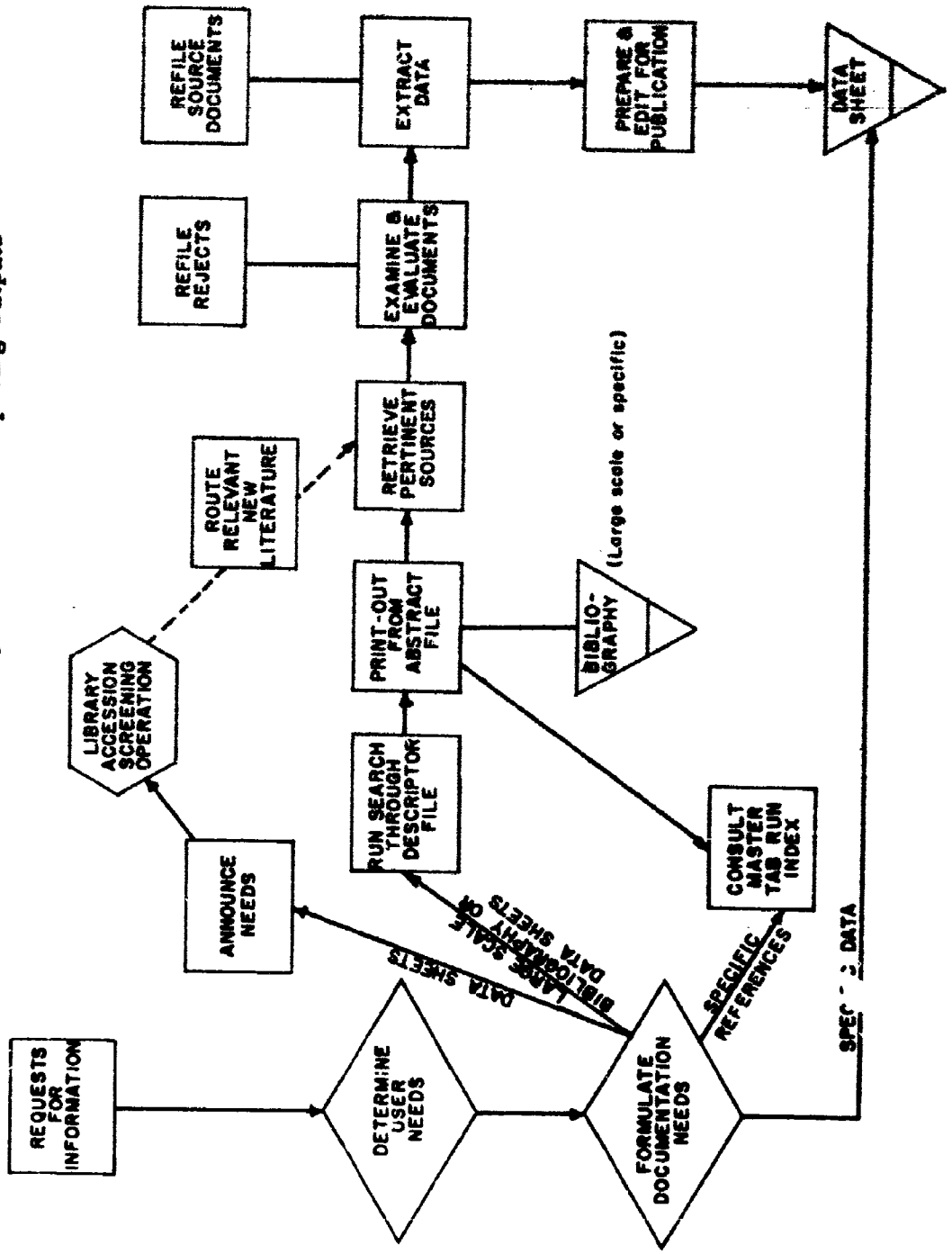
EPIC #1

Construction and Maintenance of Data Source File



EPIC #2

Searching of Files and Preparing Outputs



Atomic and Molecular Processes Information Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

The Atomic and Molecular Processes Information Center evaluates, reviews and retrieves data on atomic and molecular processes. Scientific interest in physical and chemical processes on an atomic scale has been expanded from the study of laboratory gas discharges to include potential thermonuclear plasmas, astrophysical systems and magneto-hydrodynamic phenomena, as well as numerous gas and plasma systems not in thermodynamic equilibrium. Collisions of a binary and, less frequently, of a ternary nature dominate the plasma state in the fourth state of matter.

Major areas of interest are the interaction of heavy particles with other heavy particles, electron and heavy particle interactions with macroscopic matter, and the effects of static or quasi-static electric and magnetic fields on heavy particles, especially those in excited states.

The Center produces bibliographies in these areas on a yearly basis and is also preparing critical reviews of published and unpublished data in the fields of ion-atom rearrangement collisions and of ionization, excitation and dissociation by heavy particles. In addition, the Center answers inquiries and serves as a clearinghouse for technical information in its subject field.

Atomic and Molecular Processes Information Center

Oak Ridge National Laboratory
Oak Ridge, Tennessee

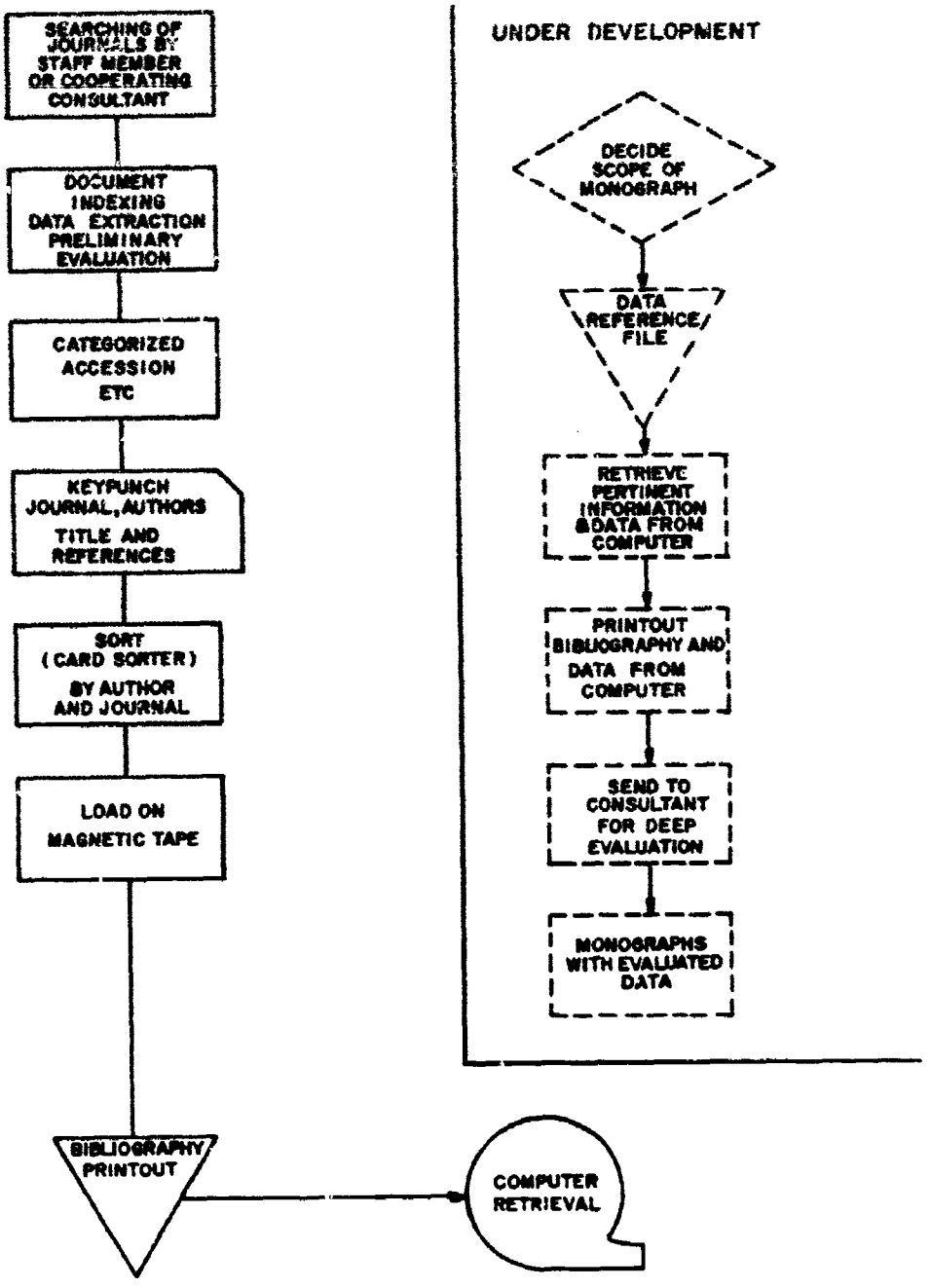
## Physical Arrangement of Data Source Files (Library)

The source document files and indices and their construction are indicated with the proper label and recording media. The following code is used to indicate each element included in the file, and its function as a mode of access for searching: P = Primary access mode; S = Secondary access mode; N = Not used as an access mode.

| | Descriptor File |
|---|---|
| Recording Media | Magnetic tape; Bibliographic file |
| File Element | |
| Source Document | |
| Document Abstracts | |
| Personal Author | S |
| Journal Source | S |
| Type of Source Document | |
| Date of Publication | S |
| Accession Number | S |
| Substance Identified | |
| Property Identified | |
| Type of Measurement | |
| "Categories" | P |
| "Terms" | S |
| Data (extracted) | |
| Title Words | S |
| Reactants | P |

Atomic and Molecular Processes Information Center

Construction and Maintenance of
Data Source File



**Science Communication**
Washington, D. C. 200 07



# APPENDIX D

## Specimen Worksheet

# WORKSHEET FOR ANALYSIS OF DATA COMPILATION ACTIVITIES

## I. OPERATIONAL CONTEXT

### A. Housing Organization

1. Name Thermal Laboratory, Dow Chemical Company, Midland, Michigan (JANAF)

2. Type of institution:
   ( ) a. Governmental
   (x) b. Industrial
   ( ) c. Educational
   ( ) d. Professional Society
   ( ) e. Other (identify) _____

3. Primary motivation for data compilation activities:
   ( ) a. In-house support servicing
   (x) b. External servicing, non-profit *(contractor operated)
   ( ) c. External servicing, profit making
   ( ) d. Other (identify) _____

4. Source of funding for data compilation activities:
   _____% a. Industry
   _____% b. Professional Societies
   100 % c. Federal Government
   _____% d. Local and State
   _____% e. Other (identify) _____

5. Percentage distribution of current data compilation efforts among the following processing functions, e.g.,

   a. 3 % for acquisition of data sources
   b. 5 % for source document screening and evaluation
   c. 2 % for source document file (library) maintenance

   d. 65 % for data screening and evaluation
   e. 5 % for data extraction
   f. 10 % for data file construction and maintenance

   } an integral reiterative process

   g. 10 % for compilation publication and dissemination

   h. 0 % for inquiry handling
   i. 0 % for other processing function(s) (Specify) _____

6. Identification of data compilations prepared by housing organization: _____
   JANAF Thermochemical Tables

Budget Bureau No. 99-R0052
Approval Expires February 28, 1968



## DATA SOURCE (LIBRARY) COLLECTION AND PROCESSING

### A. Identification and Acquisition of Sources of Data

1. Which of the following are screened in searching for candidate document titles or data sources?

   (x) a. Journal articles, Approx. % 5 .
   ( ) b. Catalog cards, Approx. % _____.
   ( ) c . Bibliographic citations in indexed documents, Approx. % _____.
   (x) d. Index and abstract publications, Approx. % 5 . *Chemical Abstracts
   (x) e. Computerized tape indexes, Approx. % 85 . * Chemical Titles
   ( ) f. Punched card indexes, Approx. % _____.
   (x) g. Others (Identify) Research Reports , Approx. % 5 .

2. Total number of candidate sources identified annually 1,500 .

3. Which of the following data sources are annually indexed?

   (x) a. Journal articles, Approx. % 90 .
   (x) b. Published reports other than journal articles, Approx. % 10 .*Govt. research reports
   ( ) c. Unpublished laboratory reports, Approx. % _____.
   ( ) d. Data sheets from other compilations, Approx. % _____.
   ( ) e. Customized data sheets, Approx. % _____.
   ( ) f. Others (Identify) _____, Approx. % _____.

4. What is the form of data sources currently accessioned?

   (x) a. Hardcopy, Approx. % 100 .
   ( ) b. Microfilm, Approx. % _____.
   ( ) c. Microfiche, Approx. % _____.
   ( ) d. Punched cards, Approx. % _____.
   ( ) e. Magnetic tapes, Approx. % _____.
   ( ) f. Others (identify)
   _____, App. % _____.
   _____, App. % _____.
   *Planning conversion to microcards .

5. What is the total number of data sources indexed to date? 26,000

6. What is the period covered by data sources indexed to date? 1925-66

7. What percentage of the total, usable data sources issued during the period of coverage has been indexed? >90 %

8. What percentage of currently issued, usable data sources is indexed? >90 %

9. What percentage of current year accessions were issued or published within the past year? 90 % *in abstract bulletins; 50% in original journal.

### B. Data Source Screening and Evaluation

1. What are principal criteria for screening of titles? The document must report measurement of property of interest to Thermal Laboratory; no substance screening is performed. Screening is done automatically by use of weighted key words.

2. What percentage of initially considered titles is rejected without physical examination of the data source? >95%

3. What are the principal criteria for screening of data sources? Does the candidate source contain new measurements. No substance screening is performed; remeasurement of a property value is accepted.

4. What percentage of data sources physically examined is rejected? 50%

C. Data Source Cataloging and Indexing

1. Indicate the extent to which the following modes of cataloging and indexing are used.

| | a. | b. | c. | d. |
|---|---|---|---|---|
| Means of Indexing or Cataloging | Total # of such descriptors used to date | Current ann. rate of addition of new descriptors. | Av. # of descriptors assigned to each source doc. | Cur. annual growth of descriptor entries |
| Substance Ident. | apprx. 10,000 | apprx. 1,000 | | |
| Property Ident. | 70 | | | |
| Type of Measrmt. | 18 | | | |
| Others (identify) | | | | |
| Substance Charac. | 9 | | Total of 5 | 7,500 entries |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

2. Indicate which means are used to index the substance coverage of data sources.

( ) a. Trade name
( ) b. Trivial name of compounds, complexes, systems, etc.
(x) c. C. A. or other systematic chemical nomenclature
( ) d. Chemical fragment codes.
( ) e. Others (identify) _____

3. What nomenclature standards, thesauri, code manuals, etc. are used to control indexing of substance coverage? Modified Hill Indexing System, Jrl. American Chemical Society, 22, 478-94 (1900) as used by Chemical Abstracts Service and the U. S. Patent Office

4. What means are used to index the property coverage of data sources? 70 property identification keywords, 18 keywords indicating measurement or data compilation techniques and methods, and 9 keywords indicating the state or condition of the substance ore indexed by memonic 3- or 4-letter codes.

D. Physical Arrangement of Data Source Files (Library)

1. To indicate how the source document file (library) and indices are constructed, label each file and give its recording media (e.g., hardcopy, microfilm, punched cards, etc.). Then use the following code to indicate for each element included in the file its function as a mode of access for searching: P = Primary access mode, S = Secondary access mode; N = Not used as an access mode.

| File Element | 1925-62 Source Doc. File Index | 1962-66 Source Doc. File Printout | Property-Substance File Index | Data Source File + Data File | Basic Data File | File |
|---|---|---|---|---|---|---|
| Recording Media | 3x5 cards | mag. tape combined file | mag. tape | hardcopy | IBM punched cards | |
| Source Doc. | | | | N | | |
| Doc. Abstracts | | | | N | | |
| Personal Auth. | X | S | | | | |
| Journal Source (Coden abbreviation) | X | S | | | | |
| Type of Source Doc. | | | | | | |
| Date of public'n. | X | S | | | | |
| Accession Number | P | P | S | | | |
| Substance Ident. | X | | P | P | P | |
| Property Ident. (comb.) | X | | P | | S | |
| Type of measrmt. | | | | | | |
| Data Extracts | | | | | N | |
| C. A. Reference | | X | | | | |
| Form of Document | (not yet implemented) | | | | | |
| | | | | | | |
| | | | | | | |

E. Purging of Data Source File (Library)

1. Indicate the method of purging data source file (library).
   (x) a. File is not purged
   ( ) b. Source documents of a specified age (e.g., 10 years) are purged
   ( ) c. Source documents are purged when more reliable sources become available.
   ( ) d. Source documents are purged when their pertinent data is incorporated into the data file.
2. How many data sources are annually purged from the file? none

F. If this data source file (library) supports more than one data file or compilation, identify the file and compilation(s). Dow Chemical Company internal thermodynamic data file

G. Flow Chart the Steps Involved in Data Source File Construction and Maintenance

(See pages 30-31.)

## III. DATA FILE CONSTRUCTION *(combined Data Source & Data File)

### A. Screening and Evaluation

1. What criteria are used to select data for inclusion in the data file?
   (x) a. None, except that the data fall within the scope of the data file and are from an apparently reliable source.
   ( ) b. Data which meets the file's criteria for substance purity, measurement reliability, etc.
   ( ) c. Critically qualified data which meets the "one best estimate" value for the property.
   ( ) d. Other (specify) _____

2. When is data screened for exclusion from the data file?
   (x) a. Prior to extraction from the source document.
   ( ) b. Subsequent to extraction from the source document.

3. What percentage of source documents screened provide data useful in data file construction? >90 %

### B. File Accumulation and Purging

1. What is the earliest published date covered by data items processed to present? 1925

2. What is the total number of data sources processed to date, i.e., what is the total number of documents or data sources from which data have been extracted? 25,000

3. What percentage of potentially useful data sources (see item II-A-6) has been screened for processing of data into the file? 100 %

4. What combinations of substances and properties are normally found in source documents?

| | |
|---|---|
| 70% Single substance | 90% Single property |
| 30% Multiple substances | 10% Multiple properties |

5. What is the current annual rate of processing data into the file?
   1,500 additional source items included
   1,000 apprx. additional substances included
   0 additional properties included
   NA new values (i.e., for substances and/or properties not previously included)
   NA additional or updated values for substances and properties previously included

6. What percentage of data extracted is in each of the following forms?
   85 % Digital or alphanumeric  5 % Formula *(equations)
   10 % Graphic  ____ % Other (identify) ____

7. What is the frequency of updating of the data file?
   (x) a. Continuously
   ( ) b. At intervals (describe interval and basis) ____

8. How are data files purged?
   ( ) a. Files are not purged. *(to data source file)
   (x) b. Property values are retired when more reliable values become available.
   ( ) c. Files are retired when a data compilation is published.
   ( ) d. Other (identify) ____

C. Chemical Substance Coverage

1. What is the intended substance coverage for this file? All chemical compounds of interest to the Dow Chemical Company

2. What is the total number, to date, of chemical substances included in this data file? approx. 10,000.

3. What is the distribution of coverage among the following groups of chemical substances?
   <5 % Chemical Elements
   ____ % Single Chemical Elements
   ____ % Elemental Mixtures & Systems (including Alloys)
   apprx. 47.5 % Inorganic Compounds, Mixtures & Systems
   ____ % Inorganic Compounds
   ____ % Inorganic Mixtures & Systems (including Solutions)
   apprx. 47.5 % Organic Compounds, Mixtures & Systems
   ____ % Organic Compounds
   ____ % Organic Mixtures & Systems (including Solutions)
   0 % Generically Identified Substances
   ____ % Polymers, Rubbers, Etc.
   ____ % Systems, Composites
   ____ % Natural Products
   ____ % Minerals
   ____ % Refractories
   ____ % Other ____
   ____ % Other ____
   ____ % Other ____
   ____ % Other ____
   ____ % Other

4. Identify the different phases, forms, concentrations etc., of substances included in the data file. Solid, liquid, gaseous

5. If this effectively increases the number of substances covered in the file, indicate the percentage increase over item C-2. 100 %

D. Property Coverage

1. What is the intended property coverage for this file? Thermochemical and related fundamental properties.
2. What is the total number, to date, of properties included in this data file? 70.
3. How many new properties are annually added to the file? 0.
4. What is the average number of values per property-substance included in the file? N.A.
5. What factors determine the number of values compiled for each property? The number of reputable values which can be located in the literature or calculated from basic data

E. If this data file supports more than one data compilation, identify the compilations. JANAF Thermochemical Tables, Dow Thermochemical Data Files

F. File Formatting

1. Is the source document format retained or is the format converted to a standardized format? (x) Format retained. 90% (x) Format converted. 10%
2. What is the current volume of the file? 8,000 File Units. N.A. Data Records.
3. Use the following table to indicate the manner of recording information in the data file or files.

   This file is an unstructured data source and working file which contains source documents, extracts, computations and JANAF data sheets for the substance. The primary entry point is chemical substance.

File designation: Data Source File

| File Element | Sequence of Properties (File Level) | Type of Value | No. of Line Entries or File Unit | Number/ File (pages) | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | *FILE IS UNSTRUCTURED | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

Science Communication
Washington, D. C. 200 07
Final Report/3-20-67

G. Flow Chart the Steps Involved in the Construction and Maintenance of the Data File.

(See pages 30-31.)

## III. DATA FILE CONSTRUCTION *(Slave File)

### A. Screening and Evaluation

1. What criteria are used to select data for inclusion in the data file?
   ( ) a. None, except that the data fall within the scope of the data file and are from an apparently reliable source.
   ( ) b. Data which meets the file's criteria for substance purity, measurement reliability, etc.
   (x) c. Critically qualified data which meets the "one best estimate" value for the property.
   ( ) d. Other (specify) _____

2. When is data screened for exclusion from the data file?
   ( ) a. Prior to extraction from the source document.
   (x) b. Subsequent to extraction from the source document.

3. What percentage of source documents screened provide data useful in data file construction? 10 %

### B. File Accumulation and Purging

1. What is the earliest published date covered by data items processed to present? 1925

2. What is the total number of data sources processed to date, i.e., what is the total number of documents or data sources from which data have been extracted? 2500

3. What percentage of potentially useful data sources (see item II-A-6) has been screened for processing of data into the file? >95 %

4. What combinations of substances and properties are normally found in source documents?

| | |
|---|---|
| 80% Single substance | 80% Single property |
| 20% Multiple substances | 20% Multiple properties |

5. What is the current annual rate of processing data into the file?
   150 additional source items included
   100 additional substances included
   0 additional properties included
   N.A. new values (i.e., for properties not previously included)
   N.A.* additional or updated values for substances and properties previously included.

*properties updated for 100 substances

6. What percentage of data extracted is in each of the following forms?
   85 % Digital or alphanumeric
   10 % Graphic
   5 % Formula
   ____ % Other (identify) _____

7. What is the frequency of updating of the data file?
   ( ) a. Continuously
   (x) b. At intervals (describe interval and basis) Quarterly

8. How are data files purged?
   ( ) a. Files are not purged.
   (x) b. Property values are purged when more reliable values become available.
   ( ) c. Files are retired when a data compilation is published.
   ( ) d. Other (identify) _____

C. Chemical Substance Coverage

1. What is the intended substance coverage for this file? Light metal compounds, including hydrides, nitrides, halides, etc.

2. What is the total number, to date, of chemical substances included in this data file? 625.

3. What is the distribution of coverage among the following groups of chemical substances?
   20 % Chemical Elements
   ____ % Single Chemical Elements
   ____ % Elemental Mixtures & Systems (including Alloys)

   80 % Inorganic Compounds, Mixtures & Systems
   ____ % Inorganic Compounds
   ____ % Inorganic Mixtures & Systems (including Solutions)

   ____ % Organic Compounds, Mixtures & Systems
   ____ % Organic Compounds
   ____ % Organic Mixtures & Systems (including Solutions)

   ____ % Generically Identified Substances
   ____ % Polymers, Rubbers, Etc.
   ____ % Systems, Composites
   ____ % Natural Products
   ____ % Minerals
   ____ % Refractories
   ____ % Other _____
   ____ % Other _____
   ____ % Other _____
   ____ % Other _____

   ____ % Other

4. Identify the different phases, forms, concentrations etc., of substances included in the data file. principally gaseous, solid and liquid

5. If this effectively increases the number of substances covered in the file, indicate the percentage increase over item C-2. 52 %

D. Property Coverage

1. What is the intended property coverage for this file? Basic thermodynamic and specrographic data required to compute JANAF tables.

2. What is the total number, to date, of properties included in this data file? 20.

3. How many new properties are annually added to the file? 0.

4. What is the average number of values per property-substance included in the file? 1.

5. What factors determine the number of values compiled for each property? number required to compute thermodynamic functions

E. If this data file supports more than one data compilation, identify the compilations. N.A.

F. File Formatting

1. Is the source document format retained or is the format converted to a standardized format? ( ) Format retained. (x) Format converted.

2. What is the current volume of the file? 950 File Units.
_____ Data Records.

3. Use the following table to indicate the manner of recording information in the data file or files.

**File designation:** JANAF Thermochemical Data File

| File Element | Sequence of Properties (File Level) | Type of Value | No. of Line Entries or File Unit | Number/ File (pages) | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| Subs. ident. | file unit or punched card | SI | Card 1 | 950 | alphanumeric | punched cards | 72 |
| code number | | " | card 2 | 950 | numeric code | " | 4 |
| phase designation | | " | " | 950 | " | | 1 |
| reference temp. | | Doc. | " | 950 | numerical | | 6 |
| temp. intervals | | " | " | 950 | " | | 3 |
| initial temp | | " | " | 950 | " | | 4 |
| number of transitions | | " | " | 950 | " | | 1 |
| type of print out | | " | " | 950 | numeric code | | 1 |
| heat of formation | | Prop. Val | card 3 | - | numerical | | 6 |
| 1st atom. num. | | Doc. | " | - | " | | 2 |
| number of atoms | | " | " | - | " | | 2 |
| second atom. # | | " | " | - | " | | 2 |
| number of atoms | | " | " | - | " | | 2 |
| transition temp. | | Prop val. | card 4 | 520 | " | | |
| enthalpy of transi. | | " | " | 520 | " | | |
| temperature | | Doc. | " | 520 | " | | |
| heat capacity | | Prop. Val | " | 520 | " | | |
| entropy | | " | " | 520 | " | | |
| enthalpy | | " | " | 520 | " | | |

Science Communication
Washington, D. C. 200 07
Final Report/3-20-67

**File designation:** JANAF Thermochemical Data File (con't)

| File Element | Sequence of Properties (File Level) | Type of Value | No. of Line Entries or File Unit | Number/ File (pages) | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| molec weight | file unit or punched cd | Subs Ident | Card 4 | 430 | numerical | punched cards | 6 |
| type of molecule | | " | " | 430 | numerical code | | 1 |
| final temp | | Doc. | " | 430 | numerical | | 4 |
| number of vibra. freq. | | Prop. Val | " | 430 | " | | 1 |
| number of elect. levels | | " | " | 430 | " | | 1 |
| rotational const. | | " | Card 5 | 430 | numerical | | 6 |
| symmetry number | | " | " | 430 | " | | 1 |
| vibrational correction | | " | " | 430 | " | | 7 |
| vibrational wave number | | " | " | 430 | " | | 6 |
| quadratic vibrational ampl. | | " | " | 430 | " | | 5 |
| moments of inertia | | " | " | 430 | " | | 13 |
| vibrational frequencies | | " | Card 6 | 270 | " | | >10 |
| electronic level | | " | Card 7 | 430 | " | | >10 |
| quantum weight | ↓ | " | " | 430 | " | | - |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

G. Flow Chart the Steps Involved in the Construction and Maintenance of the Data File.

(See pages 30-31.)

**ience Communication**
**shington, D. C.**
nal Report/3-20-67

USER SERVICING

A. Service Philosophies and Emphases

1. How are the following service features rated in your current operations? Use the following scale, or comment:

P = A primary objective
S = A secondary objective
D = Desirable, but subordinated by other factors
N = Not a pertinent objective
U = Undesirable objective

(N) a. Collect and disseminate all data in area covered.
(P) b. Collect all data and disseminate only selected data in area covered.
(N) c. Collect and disseminate only selected data (e.g., most probable values) in area covered.

(S) d. Provide user with references to sources of data.
(N) e. Provide user with data usable and with reference to information sources.
(P) f. Provide user with data usable without reference to information sources.

(D) g. Provision of quick response to individual queries.
(P) h. Periodic dissemination of accumulated information or data.

(D) i. Fast processing of newly available data.
(S) j. Fast dissemination of newly available data.

(P) k. Reduce costs for user to obtained references and/or data.
(S) l. Increase utility of data by conversion to a more useful format (e.g., graphical presentations, or ADP forms such as punched cards.

2. Who are the users of the products of this data compilation activity?
(x) a. Basic Scientists 25 %
(x) b. Applied Technologists 75%
( ) c. Others _____% (Specify) _____
(x) d. Specific technical communities (Please describe: if their needs have influenced your service criteria, please comment:) _____
rocket propulsion specialists

( ) e. Can you characterize the users more definitively, e.g., how they use the data, the estimated (1) active, and (2) potential size of the community or communities, etc? (for each compilation prepared):
The JANAF Thermochemical Tables serve as the authority for thermochemical data for use in rocket propellant performance calculations. The Tables have been distributed internationally and serve a broad spectrum of basic scientists and applied technologists.

3. Activity's current processing time from data source identification to user availability.

a. N.A. for bibliographic referencing in response to query.
b. N.A. for processed data in response to an inquiry.
c. N.A. for appearance in accession lists or bibliographies.
d. Averages 6 months for distribution in data compilation (specify format):

## B. Types of Output Prepared

1. Indicate the distribution of volume of output:
_____% Data source reference without data extracts.
_____% Data extracts without data source references.
100% Data extracts and data source references.
_____% Other (identify) _____

2. Indicate the distribution of output according to the type of requirement:
_____% Generated in response to inquiries.
100% Generated for periodic publication.
_____% Other (Specify) _____

## C. Usage of Data Source Files (Library)

1. Indicate the frequency of usage of the data source file (library).

| Type of File Search | Number of Searches/Year |
|---|---|
| Searches in response to external queries for document references. | None |
| Periodic searches for preparation or updating of a bibliography or accession list for external distribution. | None |
| Searches by staff to identify sources of data required to answer data queries. | None |
| Searches by staff to identify sources of data required to prepare or update data compilations | 4 mag. tape searches; 100 manual searches of file folders. |

2. Provide the following information for accession lists or bibliographies prepared and disseminated. NONE PUBLISHED.

| Accession List or Bibliography | Frequency of Publication or Updating | Forms Distributed | Type of Indexing |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

3. What criteria are used to select data sources for inclusion in bibliographies?
NONE PUBLISHED.

4. Is the format of the data source file (library) normally retained or modified for bibliography publication? N.A.
( ) a. Retained ( ) b. Modified

D. Usage of Data File

| Type of File Search | Number of Searches/Year |
|---|---|
| Searches to obtain data required to answer the data queries. | 0 |
| Searches to obtain data required to prepare or update data compilations | Data file is not used in this manner. It is part of a computation scheme. |

E. Publication Statistics

| Data Compilation | Initial Date of Publicn | Current Tot. Vol. of Compilation | Frequency of Supplements or Updatings | Forms Distrubuted | # Copies Distributed |
|---|---|---|---|---|---|
| JANAF Thermochemical Tables | 1960 | 950 Data Sheets | Quarterly | Hardcopy | 200* |
| | | | | Mag. tapes | 20 |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

*1,364 copies of JANAF Thermochemical Tables and 863 copies of the supplement have been sold by CFSTI. Dow, under its contract, is required to produce only 200 copies of the compilation.

Provide the Following Information for Each Compilation Where the Information is Different From the Data File

Identity of Compilation JANAF Thermochemical Tables

F. Screening and Evaluation

1. What criteria are used to select data for inclusion in the compilation?
   ( ) a. None, except that the data fall within the coverage of the data compilation and are from an apparently reliable source.
   ( ) b. Data which meet the compilation's criteria for substance purity, measurement reliability, etc.
   (x) c. Critically qualified "one best value".
   ( ) d. Other (identify) _____

2. What percentage of data pulled from the data file is normally accepted for inclusion in the data compilation? 100 % *(part of table computation)

3. What percentage of potentially useful data items in the data file have been considered for inclusion in the data compilation? >95 %

4. What is the current annual rate of issuing data in compilations?
   150 additional source items included
   100 additional substances included
   0 additional properties included
   0 new values (i.e., for substances and/or properties not previously included)
   100 additional or updated values for substances and properties previously included
   _____ other additions or changes (please identify) _____

G. Chemical Substance Coverage

1. What is the intended substance coverage of this compilation? Light elements and their oxides, fluorides, chlorides, nitrides, carbides, hydrides, etc.

2. What is the total number, to date, of chemical substances included in this data compilation? 625 (estimate)

3. What is the distribution of coverage among the following groups of chemical substances?

10 % Chemical Elements
_____% Single Chemical Elements
_____% Elemental Mixtures & Systems (including Alloys)

90 % Inorganic Compounds, Mixtures & Systems
_____% Inorganic Compounds
_____% Inorganic Mixtures & Systems (including Solutions)

_____% Organic Compounds, Mixtures & Systems
_____% Organic Compounds
_____% Organic Mixtures & Systems (including Solutions)

_____% Generically Identified Substances
_____% Polymers, Rubbers, Etc.
_____% Systems, Composites
_____% Natural Products
_____% Minerals
_____% Refractories
_____% Other _____
_____% Other _____
_____% Other _____
_____% Other _____

_____% Other

4. Identify the different Phases, forms, concentrations, etc. of substances included in the data file. Solid, liquid, gaseous

5. If this effectively increases the number of substances in III-G-2, give the percentage increase over item C-2. 52 %

## H. Property Coverage

1. What is the intended property coverage for this compilation? Heat capacity, entropy, free energy function, enthalpy, heat of formation, equilibrium constant

2. What is the total number, to date, of properties included in this compilation?
6

*(all phases)

3. What is the average number of values per property-substance included in the compilation? 61.

4. What factors determine the number of values published for each property?
A value is published for each 100°C Temperature interval between 0°K and 6000°K; when a substance experiences a phase change over this temperature range.

## I. Compilation Formatting

1. Is the data file format retained or is the format converted to a different publication or dissemination format?
( ) a. Data file format retained (x) b. Format converted

2. What is the current volume of the data compilation?
950 data sheets, etc.
_____ information units

3. Use the following table to indicate the manner used to publish or disseminate the data compilation.

**Output Document Designation:** JANAF Thermochemical Data (p. 1 of 5)

| Information Unit Element | Document Level and Sequence | Type of Value | Number / Document Page | Number / Report Document | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| Temperature | FR | Doc. | 61 | 38,000 | numerical | hardcopy | 4 |
| Heat capacity | FU | Prop. Val | 61 | | " | | 5 |
| Entropy | " | " | 61 | | " | | 6 |
| Free energy | " | " | 61 | | " | | 6 |
| Chge in enthalpy | " | " | 61 | | " | | av. 6 |
| Heat of formation | " | " | 61 | | " | | av. |
| Chge in free energy | " | " | 61 | | " | | av. 7 |
| Equilib const. | " | " | 61 | | " | | av. 6 |
| CA name | doc. page | Subs. Iden | 1 | 950 | alphabetical | | av. 17 |
| molec form | " | | 1 | 950 | " | | av. 5 |
| physic. state | " | " | 1 | 950 | " | | av. 8 |
| molec. wt. | " | Doc. | 1 | 950 | " | | av. 8 |
| molec. wt. | " | Subs. Iden | 1 | 950 | numerical | | av. 5 |
| cal. $md.^{-1}deg^{-1}$ | " | Doc. | 1 | 950 | alphanumer. | | 15 |
| K cal. $mole^{-1}$ | " | " | 1 | 950 | " | | 11 |
| T. $^{o}K$ | " | " | 1 | 950 | alphabetical | | 3 |
| $C^{o}p$ | " | " | 1 | 950 | " | | 3 |
| $S^{o}$ | " | " | 1 | 950 | " | | 2 |
| $-(F-H^{o}_{25.})$ IT | " | " | 1 | 950 | alphanumerical | | 12 |

Science Communication
(Washington, D.C. 200 07
Final Report/3-20-67
NSF C-478
-D-xxv-
Budget Bureau No. 68-R823
Approval Expires February 28, 1968

Output Document Designation: JANAF Thermochemical Data ( p. 5 of 5)

| Information Unit Element | Document Level and Sequence | Type of Value | Number / Document Page | Number / Report Document | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| Revision Date | Doc. Page | Doc. | 1 | 350 | alphanumerical | hardcopy | av 15 |
| $\Delta H_{fo}$ | | | | | " | | |
| Heat of formation, $CO^oK$ | | | | | numerical | | av 7 |
| K cal, $mole^{-1}$ | | | | | alphanumerical | | 11 |
| $T_+$ | | | | < 50 | alphabetical | | 2 |
| transition temp | | | | " | numerical | | av 3 |
| $^oK$ | | | | " | alphabetical | | 2 |
| $T_m$ | | | | max 265 | " | | 2 |
| melting temp. | | | | | numerical | | av 3 |
| $^oK$ | | | | | alphabetical | | 2 |
| $\Delta H^o_{298.15}$ | | | | | alphanumerical | | 11 |
| heat of formation, $C_{298.15}{}^oK$ | | | | | numerical | | av 8 |
| error est. | | | | | " | | av. 6 |
| Kcal $mole^{-1}$ | | | | | alphanumerical | | 12 |
| $S^o$ 298.15 | | | | | " | | 8 |
| entropy C298.15 $^oK$ | | | | | numerical | | av 5 |
| deg. $^{-1}$ $mole^{-1}$ | | | | | alphanumerical | | 13 |
| $\Delta H_m$ | | | | | alphabetical | | 3 |
| heat of melting | | | | | numerical | | av 4 |

Science Communication
Washington, D.C. 200 07
Final Report/3-20-67

NSF C-47R
-D-xxvi-

**Output Document Designation:** JANAF Thermochemical Data (p. 3 of 5)

| Information Unit Element | Document Level and Sequence | Type of Value | Number / Document Page | Number / Report Document | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| Cal. Mole$^{-1}$ | doc. page | Doc. | 1 | max 265 | alphanumerical | hardcopy | 11 |
| $T_b$ | | | | | alphabetical | | 2 |
| boiling point | | | | max 153 | numerical | | av 3 |
| $^oK$ | | | | | alphabetical | | 2 |
| $\Delta H_v$ | | | | | | | 3 |
| Heat of vaporization | | | | | numerical | | av 5 |
| K cal, mole$^{-1}$ | | | | | alphanumerical | | 12 |
| $T_s$ | | | | < 50 | alphabetical | | 2 |
| sublimation temp. | | | | | numerical | | av 4 |
| deg. K | | | | | alphabetical | | 2 |
| $\Delta H_s$ | | | | | | | 3 |
| Heat of sublimation | | | | | numerical | | 6 |
| K cal, Mole$^{-1}$ | | | | | alphanumerical | | 12 |
| point group | | | | max 532 | alphabetical | | av 3 |
| ground state configuration | | | | | alphanumerical | | av 5 |
| F | | | | | alphabetical | | 1 |
| electronic wave number | | | | | numerical | | 1 |
| cm$^{-1}$ | | | | | alphanumerical | | 4 |
| g | | | | | alphabetical | | 2 |

Science Communication
Washington, D.C. 20037
Final Report/3-20-67
NSF C-478
-D-xxvii-

Output Document Designation: JANAF Thermochemical data (p. 4 of 5)

| Information Unit Element | Document Level and Sequence | Type of Value | Number / Document Page | Number / Report Document | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| multiplicity of electronic states | doc. page | doc. | 1 | max 532 | | hardcopy | 1 |
| W | | | 0 to 3 median 1 | | | | 2 |
| vibrational wave number | | | 0 to 9 median 3 | | | | av. 5 |
| $cm^{-1}$ | | | 0 to 3 median 1 | | | | 4 |
| $W_x$ | | | 1 | | | | 2 |
| quadratic vibrational amplitude | | | | | | | av. 5 |
| $cm^{-1}$ | | | | | | | 4 |
| $W_y$ | | | | | | | 2 |
| cubic vibrational amplitude | | | | | | | 5 |
| $cm^{-1}$ | | | | | | | 4 |
| $D_o$ | | | | | | | 2 |
| heat of dissociation | | | | | | | - |
| K cal $Mole^{-1}$ | | | | | | | 11 |
| $D_o$ | | | | | | | 2 |
| rotational-vibrational interaction constant | | | | | | | - |
| $cm^{-1}$ | | | | | | | 4 |
| $\beta$ | | | | | | | 1 |
| rotational constant | | | | | | | 6 |
| $cm^{-1}$ | | | | | | | 4 |

Science Communication
Washington, D. C. 200 07
Final Report/3-20-67
NSF C-478
-D-xxviii-

Output Document Designation: JANAF Thermochemical Data (p of 5)

| Information Unit Element | Document Level and Sequence | Type of Value | Number / Document Page | Number / Report Document | Type of Representation | Recording Media | Length of Element (Characters) |
|---|---|---|---|---|---|---|---|
| σ | Doc page | Sub. Iden | 1 | max 532 | alphabetical | hardcopy | 1 |
| vibrational correction | | Doc. | 1 | | numerical | | 7 |
| $cm^{-1}$ | | Prop val | 1 | | alphanumerical | | 4 |
| r | | SI | 1 to 3 median=1 | | alphabetical | | 1 |
| internuclear distance | | Doc. | | | numerical | | av 4 |
| $A^o$ | | PV | | | alphabetical | | 2 |
| bond angle | | SI | | | numerical | | av 7 |
| deg. ($^o$) | | Doc. | | | alphabetical | | 1 |
| I | | SI | 1 to 4 median 2 | | alphabetical | | 2 |
| moments of inertia | | Doc | | | numerical | | av. 13 |
| g. cm. | | SI | | | alphabetical | | 6 |
| 6 | | Doc | 1 | | alphabetical | | 1 |
| symmetry number | | SI | 1 | ✓ | numerical | | 1 |
| references | | SI | 1 to 4 median 2 | | alphanumerical | | > 50 |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

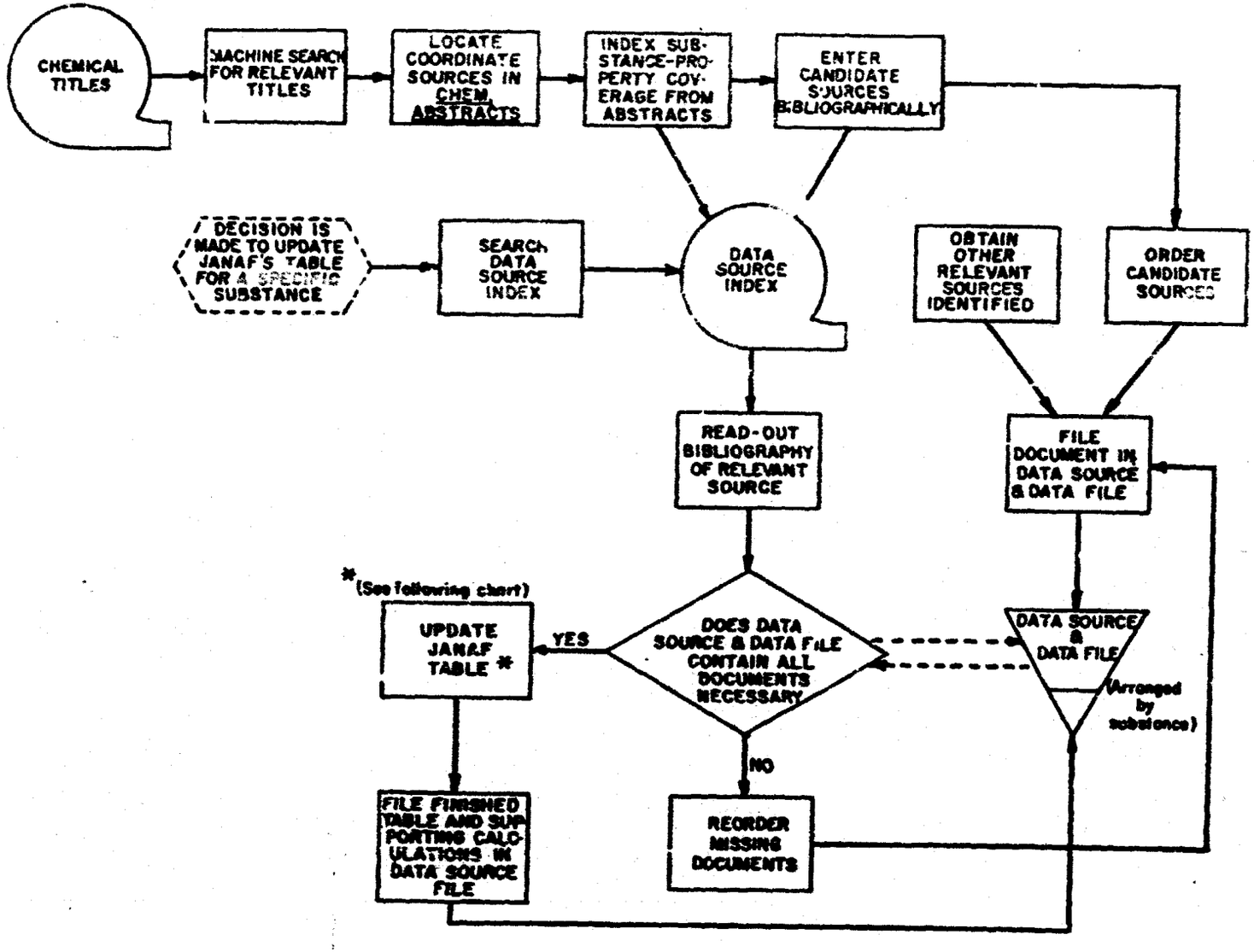
At the document level, 5 basic constants, 6 derived constants, 4 defined constants and 8 auxilary constants are defined in terms of name, symbol, value, units, and uncertainty

Science Communication
Washington, D C 200 07
Final Report 3-20-67

NSF C-475
-D-xxix-

Budget Bureau No. 48-R017
Approval Expires February 28, 1967

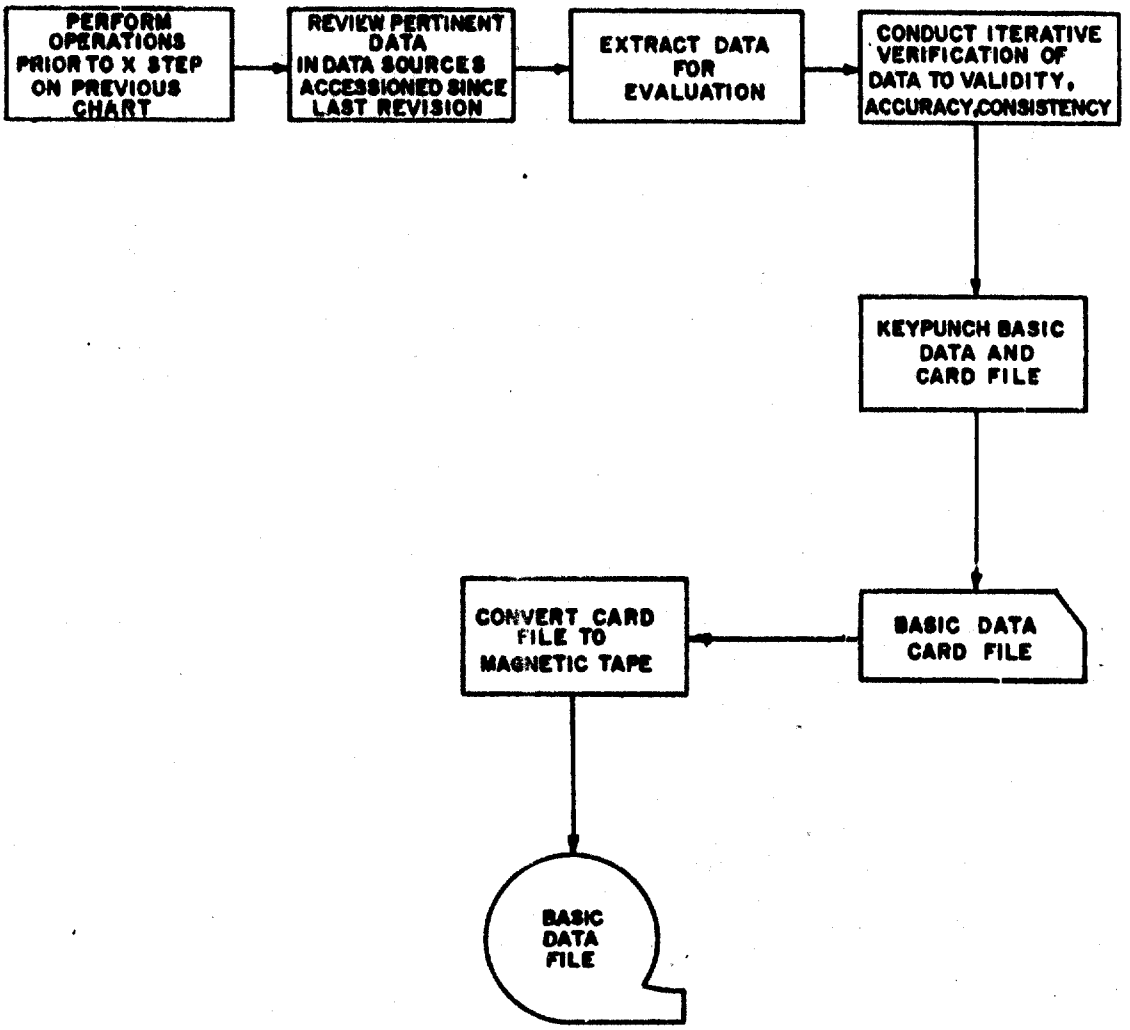
Construction and Maintenance of Data Source and Data File




Science Communication
Washington, D. C.
200 07
Contract Final Report
20 March 1967
-D-xxx-

## JANAF #2

### Construction and Maintenance of Basic Data File



An extensive, iterative computing routine is employed to compute the tables of thermodynamic functions. The input consists of the constants and fundamental data stored in the basic data magnetic tape file and a stored program for processing this data. Preparation of a detailed flow chart of the computational procedure does not fall within the scope of our current effort.

The final tables are prepared by typing and photo-reproduction from the computer printouts. Recently, a successful trial was made, jointly with the National Bureau of Standards, to automatically compose and print the data sheets.

Budget Bureau No. 99-60017
Approval Expires February 28, 1967



## V. GENERAL INFORMATION

A. Describe significant historical changes in subject scope, users served, services offered, operational mechanics, revenue sources, etc.

The JANAF Thermochemical Tables activity has achieved a stable level of operation. Future changes are expected to represent refinement of methods and introduction of improved processing techniques. The objectives, scope, etc., of the compilation activity appear stable for the near future.

B. Indicate present and potential opportunities and problems related to recent or future operations of this activity, especially as it relates to use of automatic data processing.

1) Recently a prototype supplement to the JANAF Thermochemical Tables was prepared by computer-aided photocomposition. This prototype was prepared with the assistance of the National Bureau of Standards (NBS). NBS equipment was used.
2) The Thermal Laboratory is currently shifting from maintaining hard copies of data sources to microcard copies.

C. Attach statistical information, such as size of document collection, system usage volume, personnel, budget experience, etc.

Current budget: $200,000 per year

Staff: 6 professionals

# APPENDIX E

## Specimen of Paper-tape Printout

# ANALYSIS OF DATA COMPILATION ACTIVITIES

## (TPRC) I. OPERATIONAL CONTEXT

### (TPRC) A. Housing Organization

1. Name: Thermophysical Properties Research Center, Purdue University, Lafayette, Indiana. (NSF C-478)

2. Type of institution: Educational

3. Primary motivation for data compilation activities: Primarily for in-house support servicing; secondarily for external servicing, non-profit.

4. Source of funding for data compilation activities: 100% federal government.

5. Percentage distribution of current data compilation efforts among the following processing functions: 35.5% for acquisition of data sources, source document screening and evaluation, and source document file maintenance; 57.5% for data extraction, screening and evaluation and for compilation publication and dissemination; 7% for inquiry handling.

## (TPRC) II. DATA SOURCE (LIBRARY) COLLECTION AND PROCESSING

### (TPRC) A. Identification and Acquisition of Sources of Data

1. The following are screened in searching for candidate document titles or data sources: 65% journal articles, 4% bibliographic citations in indexed documents, 30% index and abstract publications, 1% others.

2. Total number of candidate sources identified annually: Approximately 9,200.

3. The following data sources are annually indexed: 76% journal articles, 24% published reports other than journal articles (journal articles may be held in the form of an abstract only, until original document can be obtained).

4. What is the form of data sources currently accessioned? 80% hardcopy, 20% microfiche, as received.

5. What is the total number of data sources indexed to date? 33,700 indexed, 40,000 accessioned

### (TPRC) B. Data Source Screening and Evaluation

1. What percentage of initially considered titles is rejected without physical examination of the data source? Not applicable - abstracts or journals scanned directly.

3. What percentage of data sources physically examined is rejected? 50% up to now, to go down with new accession-screening procedure of direct journal scanning.

(TPRC) C. Data Source Cataloging and Indexing

1. Indicate the extent to which the following modes of cataloging and indexing are used:

   a. 45,116 substance, 7 property, 7 treatment of subject, 9 substance state, and 9 language of article descriptors are used to date.

   b. Rate of addition of substance descriptors is not available, property, subject treatment, substance state, and language descriptors are fixed in number.

   c. An average of 2 1/2 to 3 substance and property descriptors assigned to each source document, 1 subject, 1 substance state and 1 language descriptor.

   d. Approximately 12,650.

2. Indicate which means are used to index the substance coverage of data sources: C.A. or other systematic chemical nomenclature, in combination with organization rules based on chemical composition and classes of substances.

(TPRC) D. Purging of Data Source File (Library)

1. How many data sources are annually purged from the file? File is not purged.

(TPRC) III. DATA FILE CONSTRUCTION

(TPRC) A. Screening and Evaluation

1. What criteria are used to select data for inclusion in the data file? Not applicable.

2. What percentage of source documents screened provide data useful in data file construction? Not applicable.

(TPRC) B. File Accumulation and Purging

1. What is the total number of data sources processed to date, i.e., what is the total number of documents or data sources from which data have been extracted? Not applicable.

2. What is the current annual rate of processing data into the file? Not applicable.

3 What percentage of data extracted is in each of the following forms? 20% digital or alphanumeric, 80% graphic data, as presented in the open literature.

(TPRC) C. Chemical Substance Coverage

1. What is the total number, to date, of chemical substances included in this data file? Not applicable.

2. What is the distribution of coverage among the following groups of chemical substances? Not applicable.

(TPRC) D. Property Coverage

1. What is the total number, to date, of properties included in this data file? Not applicable.

2. What is the average number of values per property-substance included in the file? Not applicable.

(TPRC) E. File Volume

1. What is the current volume of the file? Not applicable.

(TPRC) IV. USER SERVICING

(TPRC) A. Service Philosophies and Emphases

1. How are the following service features rated in your current operations? Use the following scale, or comment:

P - A primary objective
S - A secondary objective
D - Desirable, but subordinated by other factors
N - Not a pertinent objective
U - Undesirable objective

(P) a. Collect all data and disseminate moderately evaluated data in area covered.
(P) b. Collect all data and disseminate highly evaluated or selected data in area covered.
(N) c. Collect and disseminate only selected data (e.g., most probable values) in area covered.
(P) d. Provide user with references to sources of data.
(P) e. Provide user with data and with information source references.
(N) f. Provide user with data without information source references.
(P) g. Provision of quick response to individual queries.
(P) h. Periodic dissemination of accumulated information or data.
(D) i. Fast processing of newly available data.
(D) j. Fast dissemination of newly available data.
(P) k. Reduce costs for user to obtain references and/or data.
(P) l. Increase utility of data by conversion to a more useful format (e.g., graphical presentations, or ADP forms such as punched cards.)

(TPRC) B. Types of Output Prepared

1. Indicate the distribution of output according to the type of requirement: Not applicable.

(TPRC) C. Usage of Data Source Files (Library)

1. Indicate the frequency of usage of the data source file (library): Daily manual searches by staff in response to external queries for document references or to identify

sources of data required to answer data queries; Periodic searches every 12 to 18 months for preparation or updating of a bibliography or accession list for external distribution (Retrieval Guide); continuous searches by staff to identify sources of data required to prepare or update data compilations.

(TPRC) D. Usage of Data File

Not applicable.

(TPRC) E. Publication Statistics

a. TPRC Data Book, 1960, 3,322, semi-annual, 11 x 17 inch looseleaf pages, approximately 400 sets.

b. Standard Reference Data monograph publications - not applicable.

(TPRC) F. Screening and Evaluation

TPRC Data Book

1. What criteria are used to select data for inclusion in the compilation? None, except that the data fall within the coverage of the data compilation and are from a reliable source; critically qualified "recommended values".

2. What percentage of data pulled from the data file is normally accepted for inclusion in the data compilation? Not applicable.

3. What is the current annual rate of issuing data in compilations? 1,391 additional source items included, 3,725 additional specimens included, 0 additional properties included, 1086 pages on new substances and/or properties included, and 642 additional or updated pages for substances and properties previously covered.

(TPRC) G. Chemical Substance Coverage

1. What is the total number, to date, of chemical substances included in this data compilation? 7,700 specimens, substances not applicable.

2. What is the distribution of coverage among the following groups of chemical substances? Not available.

(TPRC) H. Property Coverage

1. What is the total number, to date, of properties included in this compilation? 5.

2. What is the average number of values per property-substance included in the compilation? Not applicable.

(TPRC) I. Compilation Format

1. What is the current volume of the data compilation? 3,322 data sheets.

# APPENDIX F

## List of Project Meetings and Contacts

**Science Communication**
Washington, D. C. 200 07



## Project Meetings and Contacts

Date: July 15

Contact: Dr. Franz Alt, Dr. Herman M. Weisman, and Dr. Alfred Weisberg, National Standard Reference Data System, National Bureau of Standards, Washington, D. C.

Purpose: Mutual project orientation, acquire documents, explore potentials for use of survey data from usage questionnaires sent to ACS mailing list.

Date: July 20

Contact: Dr. Guy Waddington, Director, Office of International Critical Tables, National Academy of Sciences - National Research Council, Washington, D. C.

Purpose: To obtain guidance as to existing chemical data compilation activities and to arrange use of the library of the Office of International Critical Tables.

Date: July 27

Contact: Miss Marion Woodruff, Administrative Assistant to Dr. Frank Carman, Technical Director, Manufacturing Chemists Association.

Purpose: Acquire information re current status of MCA and data compilation projects.

Date: July 28

Contact: Dr. Y. S. Touloukian, Director, Thermophysical Properties Research Center, Purdue University, Lafayette, Indiana.

Purpose: Test draft version of interview guide for obtaining information from directors of data compilation activities.

Date: July 29

Contact: Mr. Harold Rienstra, Information Specialist, Systems Development Corporation, Dayton, Ohio.

Purpose: To discuss experiences in survey and analysis techniques.

Date: August 1

Contact: Mr. Kenneth Zabriskie, Chemical Abstract Services, Columbus, Ohio.

Purpose: To coordinate our project effort with the effort conducted by the Chemical Abstract Services under NSF support. Primary topics discussed included the registry system and its potential interface with systems for handling data on the properties of chemical substances and systems.

Date: August 13

Contact: Dr. Herman Weisman, Office of Information Services, Standard Reference Data System, National Bureau of Standards, Washington, D. C.

Purpose: Obtain preliminary results of survey of technical data usage by members of the American Chemical Society and to arrange for use of the NSRDS library.

Date: September 13

Contact: Mr. Daniel N. Lapedes, Managing Editor, Encyclopedia of Science and Technology, McGraw-Hill, Inc.

Purpose: Discuss data format concepts, topical organization experience in encyclopedia development.

Date: September 12-14

Contact: American Chemical Society Meeting, New York, New York.

Purpose: Attend sessions concerning chemical data.

Date: September 14

Contact: Dr. Howard Bond. Department of Chemistry, University of Rhode Island.

Purpose: Secure advice on data compilation activities in toxicology field that are desirable candidates for the project analysis.

Date: September 15

Contact: Dr. Bruno J. Zwolinski, Chemical Thermodynamic Properties Center, Texas A & M, College Station, Texas.

Purpose: Test interview questionnaire draft.

Date: October 3

Contact: Mr. Emil Schafer, Director, and Mr. John Milek, Staff Member, Electronic Properties Information Center, Hughes Aircraft Company, Culver City, California.

Purpose: Acquire specimen materials.

Date: October 4-6

Contact: American Documentation Institute Meeting, Santa Monica, California.

Purpose: Attend sessions related to project objectives.

Date: October 7

Contact: Mr. John Milek, Electronic Properties Information Center, Hughes Aircraft Company, Culver City, California.

Purpose: Discuss data format concepts, acquire specimen materials.

Date: October 10

Contact: Dr. R. E. Hultgren, Department of Metallurgy, University of California, Berkeley, California; Dr. K. K. Kelley, Retired (Bureau of Mines, Department of Metallurgy), University of California, Berkeley, California; and Mr. Brewer, Department of Chemistry, University of California, Berkeley, California.

Purpose: Acquire specimen materials.

Date: December 14

Contact: Dr. S. Rossmassler, NSRDS, National Bureau of Standards, Gaithersburg, Maryland

Purpose: Discussion and exploration of specific data center activities.

Date: December 28

Contact: Dr. L. E. Kuentzel, Wyandotte Chemicals Corporation, Wyandotte, Michigan.

Purpose: Acquire specimen materials and discussion of activities.

Date: December 29

Contact: Dr. Roger Simard, Manager, X-Ray Powder Diffraction Data System, ASTM, 1916 Race Street, Philadelphia, Pennsylvania.

Purpose: Acquire specimen materials and discussion of activities.

Date: January 17

Contact: Dr. V. J. Johnson, Chief, Cryogenic Data Center, National Bureau of Standards, Boulder, Colorado.

Purpose: Acquire specimen materials and discussion of activities.

Date: January 18

Contact: Dr. Ralph Hultgren, Department of Metallurgy, University of California, Berkeley, California.

Purpose: Acquire specimen materials and discussion of activities.

Date: January 20

Contact: Mr. Emil Schafer, Chief, Electronic Properties Information Center, Hughes Aircraft Company, Culver City, California.

Purpose: Acquire specimen materials and discussion of activities.

Date: January 23

Contact: Dr. C. F. Barnett, Chief, Atomic and Molecular Processes Data Center, Oak Ridge National Laboratories, Oak Ridge, Tennessee.

Purpose: Acquire specimen materials and discussion of activities.

Date: January 24-25

Contact: Dr. Bruno J. Zwolinski, Chemical Thermodynamic Properties Center, Texas A & M, College Station, Texas.

**Science Communication**
Washington, D. C. 200 07



Purpose: Acquire specimen materials and discussion of activities.

Date: January 31

Contact: National Science Foundation Symposium on the Chemical Abstract Services, Washington, D. C.

Purpos : Attend conference for additional information on related aspects of project objectives.

Date: February 2

Contact: Mr. D. Wagman, Chemical Thermodynamic Data Group, National Bureau of Standards, Gaithersburg, Maryland.

Purpose: Acquire specimen materials and discussion of activities.

Date: February 2

Contact: Dr. D. Stull, JANAF, Dow Chemical Company, Midland, Michigan.

Purpose: Acquire specimen materials and discussion of activities.

Date: February 13

Contact: Mr. D. Wagman, Chemical Thermodynamic Data Group, National Bureau of Standards, Gaithersburg, Maryland.

Purpose: Final discussion of activities.

Date: February 14

Contact: Dr. G. Waddington, Director, Office of International Critical Tables, National Academy of Sciences - National Research Council, Washington, D. C.

Purpose: To obtain further perspective on critical data activities being investigated.

Date: March 6

Contact: Dr. Y. S. Touloukian, Director, and associates, Thermophysical Properties Research Center, Purdue University, Lafayette, Indiana.

Purpose: To finalize investigation of activity.